# TRENDS: PC-SOFTWARE, USERS GUIDE AND DOCUMENTATION FOR ROBUST GRAPHICAL TIME SERIES ANALYSIS OF LONG TERM SURFACE WATER QUALITY RECORDS

SEPTEMBER 1991

Ontario
Environment
Environnement

ISBN 0-7729-8779-3

# TRENDS: PC-SOFTWARE, USERS GUIDE AND DOCUMENTATION FOR ROBUST GRAPHICAL TIME SERIES ANANLYSIS OF LONG TERM SURFACE WATER QUALITY RECORDS

Report prepared by:
Byron A. Bodo
Water Resources Branch
Ontario Ministry of the Environment

SEPTEMBER 1991

Cette publication technique
n'est disponible qu'en anglais.

Copyright: Queen's Printer for Ontario, 1991
This publication may be reproduced for non-commercial purposes
with appropriate attribution.

PIBS 1679
log 91-2309-064

## TRENDS: PC-Software, Users Manual and Documentation for Robust Graphical Time Series Analysis of Long Term Surface Water Quality Records

TRENDS are a set of three PC computer programs named TRX, TR1 and TR2, for performing graphical time series analyses of temporal trends in long term surface water quality data series. The programs were developed to handle the unique features of surface water quality sampling records, specifically irregularly spaced series of virtually instantaneous observations, that are not readily amenable to conventional statistical methods developed for regular time series. The programs generate laser printer graphical images that are readily interpretable by individuals with minimal statistical training. Hard disk report files are also produced for more conscientious analysts seeking deeper understanding of the particular data series and the underlying dynamical processes.

### 0.0 HARDWARE REQUIREMENTS:

**The current edition of the TRENDS programs is designed for implementation on a 386 IBM PC or PC–clone with a 80387 math coprocessor operating under DOS 3.31 or higher and connected to an HP LaserJet Series II laser printer.**

**The laser printer should have at least 4 megabytes of extended memory in order to hold the graphics images produced by TRENDS programs.**

### 1.0 GRAPHICAL ANALYSIS:

Although the time series models described subsequently are numerically founded, the TRENDS software applies the results primarily to the production of graphics that are intended to provide users with a visual synopsis of the historical surface water quality record under consideration and for inclusion in written reports. Programs also generate a report file containing numerical summaries to assist interpretation of the graphics. This first version of TRENDS programs does not support the formal statistical tests mentioned further on. Integration formal testing with the numerical/graphical procedures of the current version is expected in future editions of the software.

Since viewers retain graphical images far longer than written verbal and tabular accounts, sound principles of visual and graphical analysis as espoused by Chambers *et al.* (1983), Tufte (1983) and Cleveland (1985) received considerable attention in the development of the TRENDS graphics. The current edition produces generally sound graphics most of the time; however, in instances, the axis scaling routines of the PLOT88 plotting software used to generate the plots, yield unsatisfactory results that cannot be overridden. When the scaling problem occurs, data are confined to a relatively small portion of the available graphing space. PLOT88 also suffers from poor screen graphics and a limited selection of output devices for generating hard copies. In future, it is expected that superior plotting software can be found to replace PLOT88.

One criticism of the software by users of the prototype versions was that it can rapidly generate prodigious quantities of paper output. However, the alternatives are exceedingly time consuming. An early interactive version suffered from poor screen graphics and the general difficulty that the decision to stop the analysis at some particular point really cannot be made until the analysis has run to completion. It is the author's practice to file the complete output and also to edit report files before printing. Graphics required for reports can be selectively regenerated as required. Computer modelling has always demanded the generation of copious hard copies. The results of trial runs can always be recycled.

## 2.0 REPORT OUTLINE:

The report is structured as follows:

- Introduction
- PART I: review of time series analysis and filtering concepts for analysis of surface water quality data
- PART II: instructions for running TRENDS programs including
  - preparation of input data files
  - preparation of program control (.CNT) files including selection of control parameters
- PART III: instructions for interpreting report files generated by TRENDS programs
- Appendix A: Figures (compiled separately)
- Appendix B: Tables (attached to main report)
- Appendix C: Example Report Files (attached to main report).

Users are advised to read the report thoroughly at least once prior to attempting to run TRENDS programs. The material in PARTs I & II does overlap somewhat and will likely be rearranged in future. Because it is unavoidable that an instructional guide to graphical software include copious numbers of example graphics (81 figures), the figures that accompany the text have been compiled separately; so that, readers may readily flip through the figures while reading the text.

## 3.0 INSTALLATION & GENERAL USAGE:

Installation is accomplished by merely copying the program diskettes onto the particular hard disk drive where the programs are to reside. It is recommended that the individual programs, i.e. TRX.EXE, TR1.EXE, TR2.EXE, be installed in their own separate directories, e.g.

- D:\TRX
- D:\TR1
- D:\TR2

Input data files should likewise be placed in separate directories.

Programs are then activated by typing the respective program name, e.g. TRX and hitting the enter key. The program will respond with

TYPE CONTROL FILE NAME:

The user then responds with the appropriate control file name, e.g.

GRNO3.CNT

say for a Grand River nitrate series. Note that if the .CNT file resides in the same directory as the program, simply entering the file name is sufficient. Otherwise, the full path name is required, e.g.

F:\GRAND\GRNO3.CNT

To work quickly it is recommended placing .CNT files in the same directory as the programs, particularly since .CNT files are specific by at least one or two lines to the program being run.

As for choosing amongst the three programs, there are no hard rules. A trial run with program TRX (say with 2 'runs' and 5 'iterations', see PART II) should indicate if seasonality is strong enough to warrant the seasonal adjustment model. If seasonality is not strong, program TR1 should be run. When there is strong seasonality present in the series, users may wish to run program TR2.

To assist users in getting acquainted with the software, several of the sample data sets used in this report have been compiled on a diskette along with the program specific .CNT files which are stored in separate directories.

# PART I: TIME SERIES ANALYSIS OF SURFACE WATER QUALITY DATA

## 1.0 MOE SURFACE WATER QUALITY MONITORING DATA:

Within the Ministry of Environment's [MOE'S] Water Resources Branch [WRB] long term surface water quality records are generated:

1. at the river and stream sampling sites of the Provincial Water Quality Monitoring Network [PWQMN].
2. for various inland lake sites administered by Limnology Section.

Responsibility for the central PWQMN data base currently rests with Watershed Management Section [WMS] and WMS staff are expected to be primary users of the software. In reality, beyond some river monitoring activities specifically administered by WMS, the PWQMN is comprised mainly of sites operated by the 6 MOE regions. Hence, the programs are intended for use by MOE Regional Office staff. Since throughout southern Ontario, many of the river basin Conservation Authorities [CA's] participate actively in the PWQMN program, there may be many CA staff for whom these programs should be of assistance.

Because of their high quality, a long term lake data set provided by Limnology Section served as one of the principal test data sets employed for software development.

Other potential WRB data series to which the programs could be applied might include long term:

1. Great Lakes & connecting channel water quality
2. drinking water intake quality
3. municipal and industrial effluent quality
4. groundwater quality.

Since the programs were developed specifically for typical river and lake quality records, applications to other types of water quality data cited above is not guaranteed. Occasionally MOE staff have recourse to data generated by other agencies, e.g. water intake records from the private sector, for which the software may also be suitable.

## 2.0 TREND ANALYSIS OF SURFACE WATER QUALITY TIME SERIES

Surface water quality sample data obtained by MOE, typically present **discrete time series of irregularly spaced quasi-instantaneous observations** . In addition, water quality time series are frequently characterized by:

- short records
- high natural variability (low signal-to-noise ratio)
- non-normal sample populations with high skew and inhomogeneous variance
- numerous extreme values or outliers
- seasonal variation
- covariate effects, e.g. flow, temperature, etc.
- changes in measurement methodology over time
- censored, i.e. < or > data, according to laboratory detection limits.

These features are frequently in violation of the underlying assumptions on which conventional statistical methods rely. Through the 1980's considerable progress has been achieved with the development of statistical

methods suitable for difficult to handle environmental data sets; however, most of these specialized techniques are not yet readily available in the common commercial statistical packages.

From a water quality perspective, time trend analyses of historical water quality sampling records are desired to ascertain if the ambient levels of constituents of concern have increased or decreased either as the result of:

- emerging problems
- remedial actions.

Time trend analysis is a particular aspect of the broader discipline of *time series analysis* that is concerned with changes over the course of time that might be observed in a series of observations. Temporal trend is a somewhat arbitrary concept dependent on:

- the time horizon of interest to the analyst, e.g. short, medium or long term shifts
- the time horizon of the available data series
- the time scale of measurements.

Two fundamental approaches have evolved for time trend analysis of difficult surface water quality time series:

- formal statistical tests for trend
- graphically oriented numerical procedures that permit the visual assessment of data series behaviour.

Hybrids of the two approaches have also been developed.

## 3.0 FORMAL STATISTICAL TREND TESTS:

The principal formal statistical tests for trend in surface water quality time series are nonparametric rank order statistics, i.e. based on data ranks, that depend to a much lesser degree on the usual statistical assumptions, e.g. underlying normal distributions, than the common $t$–test. These tests include several variations of the seasonal Mann-Kendall test (Hirsch *et al.*, 1982; Hirsch and Slack, 1984; van Belle and Hughes, 1984) which is functionally analogous to determining Kendall's nonparametric rank correlation coefficient $\tau$ between water quality concentrations and time. More recently McLeod and Hipel (1990) have developed a test based on Spearman's nonparametric partial $\rho$ statistic that performs at least as well as the SMK tests. While superior to parametric tests, e.g. $t$–test, experience shows that these tests are not perfect and that several alternative procedures should be applied jointly.

It is important to note that formal statistical tests are designed to assess the net or global monotonic trend over the time horizon of the data series under investigation (note that monotone trend is always in the same direction or equivalently there are no reversals of direction). While these tests are useful for assessing net long term change in a data record, they will not reveal short term, e.g. year-to-year variations, that in many situations may be of equal or greater concern to the analyst. The formal test statistics also are also rather abstract and tend to deter analysts with limited statistical training. For these reasons, the graphically oriented numerical methods of the TRENDS software were developed.

## 4.0 TIME SERIES MODELS

To develop numerical time series models, it is necessary to formulate a conceptual model of the essential features of the data. Let $Z_i$ represent be a series of water quality concentrations obtained at chronologic times $t_i$. Consider $X_i$ to be a transformed series, e.g. $Z_i = \log_{10} X_i$ where the family of permissible transforms includes the null transform, i.e. $Z_i = X_i$. In the discussion below, the various model components are developed in an algebraic sense. Details of fitting procedures are outlined in the following section. For the present, visualize both the trend and seasonal components described below as smooth lines drawn through noisy data clouds as in Figures 1a and 1b. Fitting procedures and graphical features are explained in the later sections.

### 4.1 Simple Model:

Program TR1 performs simple trend analysis.

For a series of noisy surface water quality measurements $X_i$ obtained at times $t_i$, the simplest conceptual model for the purpose of trend analysis is:

$$X_i = T_i + \varepsilon_i \tag{1}$$

where $T_i$ is the trend component and $\varepsilon_i$ is an irregular residual noise term. Usually in time series analysis, 'trend' $T_i$ is considered to be some measure of central tendency and it is the changes in $T_i$ over time that concern us.

### 4.2 Seasonal Adjustment Model:

Programs TRX and TR2 perform seasonally adjusted trend analysis.

Many water quality variables are known to exhibit regular annual periodic variations or seasonal cycles in which case it is necessary to consider the more complex **seasonal adjustment** model:

$$X_i = T_i + S_i + \varepsilon_i \tag{2}$$

which includes the seasonal term $S_i$. Models of this general conceptual form are used to produce the myriad of 'seasonally adjusted' economics, unemployment, housing starts and other statistics regularly reported in the news media.

From the perspective of time trend analysis, the seasonal component is a nuisance factor that should be eliminated in order to obtain reliable results. From a broader scientific perspective, information about seasonal variation can provide considerable insight into the underlying dynamic processes governing the behaviour of water quality variables, and thus lead to better conceptual models of water quality processes and also better monitoring and survey design.

Suppose for a moment that the trend term $T_i$ was constant over the horizon of the data series, say at the mean level of the series, i.e. all $T_i = \overline{T} = \sum_1^n X_i/n$. If strong seasonality, say a sharp rise over the summer months, were present in the data series, we should be able to observe its presence by reordering the residuals $X_i - T_i$ or, since $T_i$ are constant, observations $X_i$ into the sequence $X_j$ according to the 'seasonal times' $\tau_j$ at which the observations occur within the year. As before we would consider defining the seasonal $S_j$ as some measure of the central tendency of observations $X_j$ occurring in the vicinity of seasonal time $\tau_j$. In this respect the definition of $S_j$ amounts to determining seasonal trend within the seasonally reordered data according to the model

$$X_j = S_j + \epsilon_j \tag{3}$$

in direct analogy to model (1). Finally we reorder $S_j$ to chronologic times $t_i$ to obtain the seasonal component $S_i$ for use with model (2).

Note that the seasonal $S_i$ as defined above is essentially a statistical measure of net systematic annually recurrent effects embedded in the data series. In individual years the seasonal effect, e.g. an annual peak, may arrive earlier or later than the net peak observed in a record. Likewise, in any given year, the magnitude of that peak may be greater or lesser than the expected net behaviour. Also without further knowledge or additional measurements we cannot immediately associate $S_i$ with specific physical, chemical, biological or anthropogenic phenomena.

#### 4.2.1 Model Fitting by Iterative Decomposition:

If trend term $T_i$ is not constant, the procedure for separating observations $X_i$ into $T_i, S_i$ and $\varepsilon_i$ rather more difficult than for the simple model (1). Typically an **iterative decomposition scheme** is applied as follows:

1. with simple model (1), develop a trial estimate of trend $T_i^1$
2. obtain **de-trended residuals** $\nu_i = X_i - T_i^1$
3. reorder de-trended residuals $\nu_i$ by time within the year to obtain the series $\nu_j$

4. from $\nu_j$ and conceptual model (3), obtain trial seasonal $S_j^1$ noting that $S_j^1$ define a relative seasonal component that is centred about 0
5. obtain $S_i^1$ by reordering $S_j^1$ to chronologic times $t_i$
6. obtain **de-seasonalized residuals** $\eta_i = X_i - S_i^1$
7. obtain a new trend estimate $T_i^2$, by applying model (1) to the de-seasonalized residuals $\eta_i$
8. repeat steps 2–7 until estimates of $T_i$ and $S_i$ stabilize.

Because the initial guesses at the trend $T_i^1$ and the seasonal $S_i^1$ may be rather crude, the iterative approach is advised. The matter is discussed more thoroughly in the section describing program control parameters. McLeod *et al.* applied an iterative decomposition scheme for the analysis of monthly water quality time series. The iterative smoothing procedures developed for the TRENDS programs are also a form of ***back fitting*** (Hastie and Tibshirani, 1990).

Once the components $T_i^1$ and $S_i^1$ have been successfully estimated, the same notions can be employed to **recompose** the series in alternative configurations. For example, one may adjust the entire series to a particular base level, e.g. $\widehat{T}$, and explore the seasonal probabilities of exceeding certain water quality criteria, as per Bodo's (1991) examination of seasonal atrazine exceedences of Canadian Water Quality Guidelines. Such features will likely become available in future editions of the software.

**N.B.** When the $\log_{10}$ transform has been invoked, to be interpretable in the natural domain of measurement, the seasonal function must be added to some chosen trend level in the transform domain before back transformation. This feature is not currently available, will likely appear in future editions.

### 4.3 Covariate & Seasonal Adjustment Model:

**The following models provide an appreciation of the limitations of current TRENDS software and of what future software development is required.**

Although models (1–3) are the basis of the TRENDS software, it is important to consider this class of models because the successful determination of trend $T_i$ against the high noise background typical of many water quality data series involves eliminating determinable systematic sources of variation to the greatest extent possible. For many water quality variables, variation is frequently induced by certain readily measurable covariables. For example, in rivers the instantaneous flow rate exerts considerable influence over sediment transport capacity; hence, variations sediment associated variables such as phosphorus are often largely attributable to flow effects. Likewise water temperature is a primary determinant of dissolved oxygen levels.

The availability of a single coincidentally observed covariate series, say flows $Q_i$, leads to the possibility of models such as:

$$X_i = T_i + C_i + S_i + \varepsilon_i \tag{4a}$$

$$X_i = T_i + f_i(C, S) + \varepsilon_i \tag{4b}$$

where $C_i$ represents a systematic covariate effect in model (4a) and $f_i(C, S)$ in model (4b) represents a joint covariate–seasonal effect that is not necessarily linearly additive as in (4a). Model (4a) is usually applied in a manner analogous to the seasonal decomposition scheme cited above, but without iterations, i.e.

1. a concentration–covariate relationship is fitted
2. and then used to determine covariate effect $C_i$ as the expected concentration for a given flow $Q_i$
3. from which **covariate-adjusted residuals** $X_i - C_i$ are determined,
4. then any remaining seasonal variation $S_i$ is determined on residual series $X_i - C_i$
5. after which formal trend testing is performed on the final residuals $X_i - C_i - S_i$.

Work with river quality series (McLeod and Hipel, 1990; Hirsch *et al.*, 1991) has shown that:

- using flow as a covariate can significantly reduce variation in the concentration series,

- to varying degrees, the seasonal variation for certain cases is largely attributable to seasonal variation in flow,
- in many other cases, after removal of flow variation, significant seasonal variation not attributable to flow remains embedded in the data series.

At this point, further work is necessary to establish the best approach to handling jointly covariate and seasonal effects. Work with PWQMN river quality series suggests that for certain water quality variables concentration–flow relationships vary continually throughout the year, i.e. seasonally. Research is needed to ascertain if the simpler linear additive model (4a) is sufficient or if a more complex expression of concentration in terms of a joint covariate–seasonal function $f_i(C, S)$ as in model (4b) is appropriate. Also, it would seem that some form of iterative decomposition scheme would lead to a better fit of either form (4a) or (4b).

In the future, once available, this class of models should be considered as they embrace the elementary forms (1–3) and can be reduced as data dictate. A major area of application is the study of river quality series that are highly reflective of non-point source (NPS) pollution which is driven by hydrometeorologic processes that are highly random in character. The random or stochastic nature of hydrologic processes induces substantial variation in the measured water quality data series which renders the detection of water quality trends due to NPS remediation measures a major challenge. A recent MOE sponsored study (McLeod and Hipel, 1990) has shown that very subtle trends can indeed be detected by a combination of formal statistical tests and flow–seasonal adjustment techniques. Further work is required to develop these results into a convenient software package.

## 5.0 FILTERING OPERATIONS:

If we consider model (1), practically the process of defining $T_i$ amounts to a smoothing or filtering operation to separate observations $X_i$ into trend estimates $T_i$ and noise terms $\varepsilon_i$. Analogously, in the seasonal adjustment model, the terms $T_i$ and $S_i$ are determined by filtering operations performed on residual series $\nu_i$ and $\eta_i$ respectively. Given the difficult characteristics of typical surface water quality time series, the smoothing filters employed in the TRENDS software were selected to be:

- suitable for irregularly spaced observations
- resistant to high-leverage data, i.e. outliers, such that extreme behaviour is absorbed into the noise terms $\varepsilon_i$
- robust against asymmetric data, i.e. non-normal, data distributions
- independent of any predetermined trend model, e.g. linear, step, monotonic, etc.
- tunable to permit the examination of short, medium or long term trend according to the degree of smoothing applied.

The fundamental forms of filter employed in TRENDS programs are considered below.

### 5.1 Annual Medians:

The annual median filter is used by the program TRX to determine the trend component of the seasonal adjustment model.

Although not usually considered as such, a crude but effective and readily implementable filtering operation that is robust against outliers and distribution shape, is to determine a constant robust central tendency measure such as the median of observations available for each calendar year. This is known as a 'bin smoother' (Hastie and Tibshirani, 1990). The result is a discrete, piecewise smooth trend $T_i$ with discontinuous step jumps at the beginning (or equivalently the end) of each year of the data series. An example is shown in Figure 2 where for plotting convenience all sample dates have been converted to decimal format, e.g. April 24 1985 = 85.311. This 'annual median filter' has the advantage of being very quick to calculate. To its detriment, the median is increasingly unstable as measure of central tendency for small ($< 10$) sample sizes with the consequence that the trend indicated for poorly sampled calendar years may be somewhat distorted or exaggerated. On Figure 2, this is true for 1981 where the indicated level, based on only 4 observations,

ranks amongst the 3 highest years. Analysis with the quasi-continuous smoothers discussed below support the impression that 1981 zinc levels may have been higher that those of the surrounding years; however, the suggested 1981 level is somewhat lower than that of Figure 2.

### 5.2 Robust Smoothing Filters:

Programs TR1 and TR2 offer an array of robust smoothing filters derived from a 'running median' smoother developed for regularly spaced time series by Tukey (1977) and a 'running line' $x$–$y$ scatterplot smoother developed by Cleveland (1979). Both the Tukey and Cleveland filters may be regarded as 'quasi-continuous' as on the graphics to be presented the filter output appears generally as a smooth line traced through a noisy data cloud. In reality, the methods yield a smoothed estimate, e.g. trend $T_i$ for concentration $X_i$ at time $t_i$, and it is the plotting procedure of connecting the points $T_i$ which produces the visual effect of a continuous trend line. The list of filters currently available in programs TR1 and TR2 are given in Table 1. The present discussion focuses on the concepts underlying the 'primitive' filters from which compound or hybrid filters were developed.

#### 5.2.1 Tukey Filters:

Tukey (1977) developed a series of robust filtering procedures for tracing smooth lines through evenly spaced data series. The filters are constructed from primitive operations including:

- hanning, a simple weighted average of three, i.e. $(X_i + 2X_{i+1} + X_{i+2})/4$
- repeated running medians of three, i.e. $X_i$ is replaced by the median of $(X_{i-1}, X_i, X_{i+1})$
- splitting, a procedure for rounding flat peaks and valleys comprising 2 or more successive identical values.

The Tukey filters are designed for evenly spaced series. The degree of smoothing is not especially powerful nor are these filters particularly well suited for the irregularly spaced water quality data. Nevertheless, the Tukey filter (Filter 8), specifically an 3RSSH filter preceded by a hann, i.e. H3RSSH, is useful for exploratory work as in Figure 3 where the recurrent spikey behaviour suggests there may be a strong seasonal component in the Arsenic series. Filter 8 is also useful as a quick filter for polishing the rough corners of the output of other filters.

Tukey's notion of repetitive filtering, i.e. filtering the output of a previous filtering operation, is an important concept exploited in the development of higher order filters available with programs TR1 and TR2. The concept of moving medians was also exploited in the development of the seasonal smoothing filter described below.

#### 5.2.2 Cleveland's LOWESS Filter:

Cleveland (1979) developed an algorithm known as LOWESS to perform *robust locally-weighted regression* for the purpose of exploring nonlinear relationships in $x$–$y$ scatterplots. LOWESS employs the standard linear least square regression model $y = \beta_0 + \beta_1 x + \epsilon$, locally on subsets of $m < n$ nearest neighbours of each of the series' $n$ data points $(X_i, t_i)$ where $t_i$ is the chronologic date. Moreover, the $m$ local data points are weighted according to their distance from $(X_i, t_i)$. After a locally-weighted regression estimate $\widehat{X}_i$ is obtained for each of the $(X_i, t_i), i = 1, n$ data points in the set, additional robustness weights are defined according to the absolute magnitude of the locally- weighted regression residuals $e_i = |X_i - \widehat{X}_i|$. These robustness weights are applied on a following iteration to minimize the influence of outliers. Note that it is the estimate $\widehat{X}_i$ that is ultimately used as the trend term $T_i$ at date $t_i$. Standard LOWESS involves 2 robustness iterations. Bodo (1989) and others (McLeod and Hipel, 1990; Hirsch *et al.*, 1991) have adapted the LOWESS algorithm for tracing smooth trend lines through noisy time series of water quality data.

As per Table 1, in the current edition of TR1 and TR2, standard LOWESS with 2 robustness iterations may be performed with Filter 1, while Filter 9 offers elementary LOWESS, i.e. locally weighted regression without robustness iterations. Note that LOWESS with 2 robustness iterations amounts to 3 filtering operations each with a locally weighted regression estimate developed at each point of the $n$ points of the data series. The LOWESS algorithm is a computer intensive procedure.

#### 5.2.2.1 LOWESS Control Parameters:

A critical problem in applying LOWESS to smoothing irregular time series is how to develop systematic rules for selecting the number of nearest neighbours $m$ in a manner that associates the choice with short, medium or long term trend definition. The solution developed for the TRENDS programs TR1 and TR2 is based requires two control parameters:

1. a typical annual sampling rate
2. a 'unit smoothing interval' or USI.

The typical annual sampling rate is multiplied by the USI to obtain the number of nearest neighbours $m$ used in LOWESS and derivative smoothing filters.

##### 5.2.2.1.1 Typical Annual Sampling Rate :

Users do not actually determine a typical annual sampling rate, but rather must choose from two possible ways of determining the typical annual sampling rate:

a. median rate = 1 / median$\{\Delta_i\}$

b. mean rate = 1 / mean$\{\Delta_i\}$

where $\Delta_i$ the sampling interval series defined as $\Delta_i = t_{i+1} - t_i; i = 1, n-1$. The median sampling interval is generally lower than the mean interval; hence, the median rate is usually higher than the mean rate. Experience with PWQMN river quality series reveals that for the vast majority of approximately monthly sampling sites, the median rate provides an estimate of typical annual sampling rate that is unaffected by a few long gaps in the record and is generally close to the specified monthly sampling objective.

The actual typical sampling rate employed by the programs TR1 and TR2 is determined by truncating the estimated rate and adding 1, e.g. if the calculated rate is 10.5 samples, the rate used by TR1 and TR2 is 10 + 1 = 11, in part for the reason cited below.

There is also a technical reason for using a slightly higher sampling rate than that which obtains most years. Suppose that the median rate is determined to be 13 samples/annum which is identically the number of nearest neighbours $m$ employed in the LOWESS algorithm. Note that $m$ includes the point at which LOWESS is being performed, i.e. $(X_i, t_i)$ where $t_i$ is the chronological date. Suppose that $(X_i, t_i)$ is the midpoint, i.e. the 7th of the nearest neighbours; such that, there are 6 points to the left and 6 points to the right. Further suppose that the data are spaced at equal intervals. Under this scenario, the two most remote points $(X_{i-6}, t_{i-6})$ and $(X_{i+6}, t_{i+6})$ are equidistant from $(X_i, t_i)$ and will be assigned local regression weights of 0, implying that the local regression really involves only the central 11 nearest neighbours. Because the outermost nearest neighbours are assigned 0 or very low local regression weights, when performing LOWESS smoothing with small $m$ as will frequently occur with 'monthly' PWQMN river records when year–to–year changes are emphasized, it is advantageous to have the few additional nearest neighbours that accrue from using the median sampling rate. For example, with 'monthly' river series the mean rate is 10–11 samples/year while the median rate is 12–13 samples/year. As a matter of interest, for the Arsenic series shown on Figure 3, the mean rate was 17.2 and the median rate was 19.2. At these levels, the effect in the trend term estimates is small.

The mean rate has been found to be the most appropriate measure of typical annual sampling rate for the type of monitoring conducted in certain lake monitoring programs for which samples are obtained only from May–October and a 5 month record gap from November–April is a regular feature. For these data sets, the median rate reflects the much shorter intervals between the May–October samples and may overestimate the typical annual rate by a factor or 2–3 or more. The analysis of one lake record with this type of monitoring strategy (see PART II: 5.4) showed the mean rate was 16.1 samples/year while the median rate was 45.6 samples/year. On calendar year basis, over the 14 years the annual number of observations ranged from 6–27 with a median 15 and a mean 15.4.

The mean rate may also be the best choice for certain river monitoring records which regularly contain winter record gaps due to periods of ice cover. Sampling rate information is printed as a feature of the input routines of programs TR1 and TR2. From a trial run, the best choice of of typical annual sampling rate can

be determined. If the median rate exceeds the mean rate by a factor of 2 or more the mean rate is the best choice. Also when the ratio of median rate to mean rate approaches 1.5, the mean rate is more likley to be the best choice. Diagnostic programs are under development to assist users with these choices.

#### 5.2.2.1.2 Unit Smoothing Interval:

The degree of smoothing, hence the focus on short, medium or long term trend, is further governed by the choice of a parameter referred to as the **unit smoothing interval** (USI). To emphasize year-to-year changes, the USI should be set to 1 year. For medium term changes, and greater smoothing, USI could be set to 2 or 3 years. It is also possible to set the USI to fractions, e.g. USI = .33 year, which can be useful for exploratory work with program TR1. Generally program TR2 is designed for examining trends with USI taken as integral multiples of 1 year.

### 5.2.3 Modified LOWESS (LOWESP) Filter:

A special version of the original LOWESS filter, dubbed LOWESP, was developed in an attempt to circumvent problems with data clusters that arise in many river and lake data sets due to certain years of intense sampling, i.e. special surveys. At each datum, LOWESP counts the number of observations, say $\hat{m}$, available within the specified USI. If $\hat{m}$ exceeds the number of nearest neighbours $m$ determined from the choice of sample rate and USI, LOWESP performs local regression with $\hat{m}$. This modification effectively increases the smoothing power through modest data clusters. It is most useful when year-to-year smoothing is of interest.

As per Table 1, LOWESP smoothing is obtained by specifying Filter 2.

### 5.2.4 Hybrid Filters:

Experience has shown that there is no universally perfect filter capable of handling all the potential data configurations presented by surface water quality data sets, particularly when year-to-year behaviour (USI=1) is of interest. Using LOWESS, LOWESP and Tukey's filters, various hybrid filters (Filters 3–7 in Table 1) have been constructed based on Tukey's notion of repetitive filtering. Filters 3 and 4 merely polish the respective standard LOWESS (Filter 1) and LOWESP (Filter 2) output with additional filtering by the respective parent algorithms without robustness iterations which yield little benefit when smoothing the output of previous smoothing operations. Filters 5–7 are somewhat ad hoc constructions that have in certain instances been found useful for unusual data configurations, particularly for obtaining a smooth transition through data clusters interspersed with irregular gaps (see conductance example below). Filters 5–7 incorporate alternate ways of smoothing the series as an evenly spaced series which to a degree circumvents problems created by irregular gaps in the actual record.

Experience suggests the following advice to users:

- For modestly sized data series ($n < 300$), the hybrid filters can improve performance through data series with modest data clusters intersperse with irregular gaps in the record.
- For very large data sets ($n > 500$), repetitive smoothing can be excessively time consuming. Use of the basic filters (Filter 1 and Filter 2) is recommended.
- For USI $\geq 2$ there may be little advantage to employing hybrid filters.

### 5.2.5 EXAMPLES: LOWESS and LOWESS derived filters:

The 9 robust smoothing filters, including the Tukey filter, currently available in programs TR1 and TR2 are listed in Table 1.

#### 5.2.5.1 Comparative Filter Performance: Year-to-Year Smoothing

The sequence of Figures 4.1–4.9 illustrate the behaviour of the 9 robust smoothing filters for a particular conductance series where the USI was set to 1 year, in order to emphasize year-to-year shifts and the median annual sampling interval was selected. For comparison, Figure 4.10 presents the trend line developed with

the stepwise mean annual filter. This conductance series exemplifies the problems encountered with data clusters interspersed by irregular gaps that was noted above.

When examining these plots consider the pre– and post–1977 behaviour of the trend lines merits scrutiny. The rather jagged behaviour of the pre–1977 trend smooth obtained with standard LOWESS is evident on Figure 4.1. Filter 2, i.e. standard LOWESP, offers slight improvement as per Figure 4.2. The repetitive smoothing of Filters 3 and 4 provide negligible improvement with the pre–1977 record, but noticeably refine the trend lines post–1977 as in Figures 4.2–4.3. Filter 5 produces a smooth trend line over the entire record as per Figure 4.5. Filter 5 output most closely matches the output of the stepwise filter shown in Figure 4.10 and esthetically produces the best result for the conductance series. Compound Filters 6 and 7 also have some difficulty with the pre–1977 records, while less than satisfactory results are obtained with Filters 8 and 9 which are really intended only for exploratory use with program TR1.

#### 5.2.5.2 Comparative Filter Performance: USI = 2 years.

The sequence of Figures 5.1–5.8 illustrates that the effects of doubling the unit smoothing interval, i.e. setting USI = 2 years, are comparable to the effects of repetitive filtering obtained with Filter 5 and USI = 1 year. When choosing USI $\geq$ 2, user should probably confine filter choices to Filters 1–4. Filter 8, the Tukey filter, always produces output identical to Figure 4.8 as it is unaffected by the choice of sampling interval and USI.

#### 5.2.5.3 Comparative Performance: Filter 4, USI = 3, 5, 10 & 16.6 years.

The sequence Figure 6.1–6.4, shows the output of Filter 4 under increasing USI of 3, 5, 10 and 16.6 years. Figures 4.4 and 5.4 should also be examined for the comparable results with USI = 1 and 2. For data series like the conductance example with considerable year–to–year movements, trend is probably best represent by choosing USI = 1 or 2. For series with minor year–to–year fluctuation and gradual long term changes, higher USI may be in order, to overcome the visually obscuring influence of marginal year–to–year fluctuations.

The nearly straight lines indicating downward trend with USI = 5 and 10 years, are graphical analogies of the results that can be expected of formal trend tests that measure net trend embedded in a series. The conductance series with its trend reversal since 1985 provides a good example of why reliance on formal tests alone should be avoided.

#### 5.2.5.4 Exploratory Analysis with TR1.

The primary uses of program TR1 are:

1. trend analysis when seasonal effects are minimal (see next section)
2. exploration of unfamiliar series.

Trend plots of modestly large series, e.g. $n \geq 100$, look much like those produced by program TR2. Figure 7.1 illustrates the usage of TR1 for exploring an unfamiliar series. Choosing USI = .5 year, the number of nearest neighbours $m$ is reduced to 7 and the recurrent oscillations in the trend line indicate that significant seasonality is likely embedded in the series.

Figure 7.2 illustrates a nonconventional usage of TR1 that might be of use in some instances. The plot shows a the result of running TR1 on a very small data series (13 observations) which was obtained during a brief period of 5 months. By setting USI = .1 year, LOWESP smoothing is performed with only 3 points, effectively leading to a 'join the dots' plot. Far more efficient software could be produced to generate such plots; however, as Figure 7.2 demonstrates, program TR1 can be adopted for this purpose when required.

### 5.3 Seasonal Smoothing Filter:

#### 5.3.1 Seasonal Time:

Conventional notions of seasons are generally linked to arbitrarily fixed quarterly or monthly time periods. In the following discussion it is best to think of seasonality in terms of annually recurrent periodic effects that might be modelled by a smooth continuous function as is often achieved by fitting the seasonal as the

sum of the significant Fourier harmonics, i.e. periodic trigonometric functions. To fit a smooth seasonal function demands an appropriate continuous representation of 'seasonal time', i.e. time within the year. Given the eccentricities of calendar dates, seasonal time was handled by the following decimal convention which considerably facilitates plotting.

As data are read into the software, the month, day and time of sampling are converted into the fraction of a complete year that has elapsed since January 1 0:00 hours (equivalently December 31 24:00 hours), e.g. .5 year = July 2 12:00 EST (Eastern Standard Time), for a 365 day year. The date conversion routine accounts for leap years. The chronological time is then given by the year plus the fraction, e.g. 85.5 = July 2 12:00, 1985. Generally MOE sample times are not recorded in EST throughout the year, but rather shift to Daylight Savings Time (DST) when appropriate. This introduces a small discrepancy in seasonal times, i.e. samples are shifted forward of EST by .000114 year during DST periods; however, the effects on the seasonal function produced by the following routines is generally imperceptible.

Table 2 gives a seasonal date conversion chart for a standard 365 day year.

### 5.3.2 Moving Seasonal Median Filter:

Seasonal smoothing is performed to obtain the seasonal term $S_t$ required in the seasonal decomposition routines of programs TRX and TR2. Seasonal smoothing is accomplished identically in both programs with a **moving seasonal median** [MSM] filter that derives from the data density plots of Chambers *et al.* (1983) and Tukey's (1977) moving median filters. Effectively the MSM filter amounts to a 'running bin' smoother. Like the robust smoothers employed to define the trend term $T_t$ in programs TR1 and TR2, the MSM filter is 'quasi–continuous' as the seasonal function is defined only at certain seasonal time points and 'joining the dots' produces the visual effect of a continuous function. Seasonal values required at other times are readily obtained by linear interpolation to good approximation.

The MSM seasonal smoother is specifically designed to:

- counter the influence of outliers
- counter the influence of seasonal sampling bias
- mimic the appearance of a smooth continuous seasonal function such as might be derived by fitting Fourier harmonics.

Systematic seasonal sampling bias arises frequently at sites where extra sampling is not evenly distributed throughout the year, e.g. river sites where extra spring runoff samples are obtained or lake sites which are heavily sampled over the summer season.

The year is divided into 100 intervals of 3.65 days or effectively 101 seasonal time points $\tau_k$ including the beginning and end of the year, i.e. $\tau_0 = \tau_{101}$. After reordering the data series by seasonal time, a window or bin of width $W$ is centred at the first point $\tau_0$. All data falling within $\tau_0 \pm W/2$ are gathered into an array. Given seasonal circularity, for $\tau_0 - W/2$, data are those at the end of the year. If the number of observations within the window $\ell$ exceed $\tilde{m}$, a parameter specified by the user, the median (all quartiles on the final iteration) is determined. When $\ell < \tilde{m}$, the window $W$ is expanded in small increments until sufficient data are obtained such that $n \geq \tilde{m}$ before the median is determined. The window is then reset to $W$ and moved forward to $\tau_1$ and the process repeated. Essentially the sequence of 101 medians are computed on overlapping data sets such that only a small fraction of the data change at each $\tau_k$. Two rounds of basic LOWESS smoothing with 10 nearest neighbours and 0 robustness iterations are applied to polish the sequence of 101 medians, once again taking advantage of the circularity of the seasonal data series. Finally the 100 distinct smoothed medians are summed and adjusted to yield a sum of 0. Experience shows that a slight adjustment is needed when seasonally biased sampling records are analyzed to assure that the resulting relative seasonal function is perfectly centred about 0.

### 5.3.3 Seasonal Smoothing Control Parameters:

The two seasonal smoothing control parameters are:

1. $\tilde{m}$ = required minimum number of data falling within a seasonal window before computation of statistics is performed

2. $h$ = number of seasons per year such that seasonal smoothing window width $W = 1/h$

The goal in selecting $\tilde{m}$ and $h$ is to achieve as sharp a definition of a smoothly varying seasonal variation as data permit as economically as possible. The selection procedure remains a somewhat ad hoc, heuristic process. Generally it is desirable to determine values that are appropriate throughout most of the year. Practice suggests that when the seasonal variability embedded in the data series is reasonably strong, a reasonable degree of latitude is possible in the selection of $\tilde{m}$ and $h$ with negligible effect on the results.

The primary purpose of parameter $\tilde{m}$ is to serve as a fail–safe mechanism that triggers seasonal window expansion to take over through any thin spots in the seasonal sampling distribution, e.g. at most PWQMN river sites, December–January is typically a poorly sampled time of the year. The mechanism assures a smooth transition of the resulting seasonal function through poorly sampled times of the year during which the resulting seasonal function is essentially a projection of the seasonal behaviour exhibited to the well sampled periods to the right and left of the poorly sampled period.

The purpose of parameter $h$ is to determine the size of the seasonal smoothing window $W = 1/h$. It is simplest to envision $h$ as the number of feasible distinct seasons that might be obtained if the $n$ of the series were spread evenly throughout the year. Practice has shown that if there are sufficient data, sharp seasonal definition is obtained by choosing $h = 12$.

As the process involves some trial and error, experiment with alternative values of $\tilde{m}$ and $h$ using program TRX for quick results, until acceptable values are found. If the suggested choices for $\tilde{m}$ are employed, the associated fail–safe mechanism and the two rounds of LOWESS smoothing generally maintain the resulting seasonal function within acceptable bounds. Diagnostic programs to obtain preliminary assessments of seasonal sampling bias and guide the selection of $\tilde{m}$ and $h$ are under preparation. For the present the following rules are suggested:

- Keeping in mind that the median grows less reliable as sample size decreases, parameter $\tilde{m}$, in general, should never be set $< 10$. With modest data series, i.e. $n = 150$–$250$, and some seasonal sampling bias present, values of $\tilde{m} = 11$–$13$ were found satisfactory for many of the examples shown in this guide.
- Begin with a value of $h = 12$ (corresponding to $365/12 \approx 30$ day seasons), as this usually yields sharp seasonal definition if sufficient data are available.
- Assuming data are evenly distributed across the year, estimate the expected number of observations as $\ell = n/h$.
- If $\ell < \tilde{m}$, decrease the number of seasons $h$ (equivalently increasing window size $W = 1/h$) until a suitable value of $\ell$ is obtained. Generally, integral values in the range $h = 4, 12$ are sufficient for most series of modest size. Even with fairly small data sets, e.g. $n = 90$, perceptible seasonal variation may be evident at $h = 4 \approx 91$ day or 3 month seasons or $h = 6 \approx 61$ day or 2 month seasons.
- For large data series, e.g. $n > 300$, users may consider $h = 17 \approx 21$ day or 3 week seasons or $h = 26 \approx 14$ day or 2 week seasons. When working with large data sets, if $\ell = n/h > \tilde{m}$, users should reset $\tilde{m} \approx \ell$.

### 5.3.4 EXAMPLES: Seasonal Smoothing:

The following series of examples were culled from test data sets and are selected to illustrate various types of seasonal trends and other points regarding seasonal functions developed by the TRENDS software.

#### 5.3.4.1 Clearwater Lake Examples:

Clearwater Lake is a small Shield lake located in the Sudbury area. The chemistry and seasonal dynamics indicated by the following examples are indicative of acidified, dilute waters.

**Figure 8.1 Seasonal trends in DIC, Clearwater L.**

- strong seasonality is suggested concordant with the exploratory results obtained with program TR1, Figure 7.1.
- a case where very strong seasonality is evident in a relatively small ($n = 91$) data set.

**Figure 8.2 Seasonal trends in conductance, Clearwater L.**

- An example of reasonably strong seasonality encountered in the conductance series that was employed in the Figure 4.1–4.10 sequence. Figure 8.2 was produced specifically as the companion for Figure 4.5; however, Filters 1–7 all produce results that are very close to Figure 8.2

**Figure 8.3 Seasonal trends in manganese, Clearwater L.**

- An example of modest seasonality that is unlikely to have adverse effects on trend analysis but may merit notice from a scientific perspective.
- An example of scaling problems with PLOT88. Notice the upper 40% of the graph space is blank, flattening the appearance of the seasonal function and artificially creating the impression that the seasonal my be less significant than it is.

**Figure 8.4 Seasonal trends in potassium, Clearwater L.**

- An example of a weak seasonal function that has little or no effect on the results of trend analysis but may yet be significant, i.e. nonrandom. When percentage variation contributed by the seasonal is <10% after outlier deletion (see PART II: 5.2), the seasonal function estimated by the software is likely insignificant.

**Figure 8.5a–d Seasonal trends in aluminum, Clearwater L: Varying seasonal smoothing control parameters.**

- Figures 8.5a–8.5c show the effects of varying seasonal smoothing control parameters $\tilde{m}$ and $h$ on an aluminum data set of modest size. The choices were:
  8.5a $\tilde{m} = 11$ and $h = 10$, i.e. $W = 36.5$ days
  8.5c $\tilde{m} = 13$ and $h = 8$, i.e. $W = 45.6$ days
  8.5d $\tilde{m} = 15$ and $h = 6$, i.e. $W = 60.8$ days
- In moving from Figure 8.5a to 8.5c, the estimated seasonal function does become smoother as more data are considered at each of the 100 points at which the seasonal is estimated.
- The seasonal data series is characterized by considerable scatter, but all results suggest that there may be net seasonality of minor significance embedded in the data series. There is probably no best choice. It is best not to focus on specific features, but rather to note general shape and degree of scatter.
- Figure 8.5d, the trend plot produced with the choices for Figure 8.5a, shows an dominant downward trend. All choices in fact yield the same trend result as would an analysis with program TR1.

**Figure 8.6a–d Seasonal trends in chloride, Clearwater L: Contrast of seasonal estimated via TRX against TR2 under constant trend.**

- Figures 8.6a–8.6b show respectively the trend and seasonal plots produced for the Clearwater L chloride series with program TRX.
- Figures 8.6c–8.6d show respectively the trend and seasonal plots produced for the Clearwater L chloride series with program TR2.
- Steady upward trend is the dominant feature of the trend plots produced by either programs; however, the seasonal plots show somewhat different characteristics.
- **Under a steady upward or or downward trend the seasonal plots produced by program TRX will distort the shape and other features of the seasonal function.** When estimating trend as a constant level throughout each year, in the case of the chloride data with steady upward trend, the level of the seasonal during the early part of the year is too low and during the latter part of the year is too high.
- **To investigate the form of seasonal characteristics under steady unidirectional trend it is desirable to employ program TR2.**
- When there is either little trend or year–to–year upward trend more or less equals year–to–year downward trend over the time horizon of a series, there should be little difference observed between the seasonal characteristics estimated by TRX and TR2.

**5.3.4.2 Moira River Arsenic Examples:**
**Varying seasonal smoothing control parameters.**

The upper reaches of the Moira River are contaminanted with Arsenic as the result of long abandonned gold mining operations. As the following sequence of figures illustrate that with a large data set, i.e. 400 observations, and strong seasonality. considerable latitude is possible in the selection of control parameters $\tilde{m}$ and $h$ with only minimal effects on the resulting seasonal. The choices were:

(9.a) $\tilde{m} = 15$ and $h = 26$, i.e. $W \approx 14$ days
(9.b) $\tilde{m} = 35$ and $h = 12$, i.e. $W \approx 30$ days
(9.c) $\tilde{m} = 70$ and $h = 6$, i.e. $W \approx 61$ days

Figures 9.a–9.c speak for themselves. Figure 9.d shows the historical trend plot generated as a companion to Figure 9.b. The choice of smoothing filter control parameters had no perceptible influence on trend plots. Figure 9.d is another example of the scaling problems occasionally encountered with PLOT88 software.

**5.3.4.3 Nitrate Seasonality Examples:**

Water quality variables nitrates, i.e. $NO_3^-$, and total nitrates, i.e. $NO_3^- + NO_2^-$, frequently show strong seasonal behaviour as in the following examples.

**Figure 10.1 Seasonal trends in total nitrates, Trent R.**

- extremely strong unimodal seasonality is evident in a very large data set.
- the Trent R represents a mixture of Shield and Lowland drainage.

**Figure 10.2 Seasonal trends in nitrates, Plastic L.**

- Plastic L is a dilute, acidified lake in central Ontario Shield area.
- strong bimodal seasonality is evident in the Plastic L inflows.

# PART II: RUNNING TRENDS PROGRAMS:

## 0.0 HARDWARE REQUIREMENTS:

**The current edition of the TRENDS programs is designed for implementation on a 386 IBM PC or PC–clone with a 80387 math coprocessor operating under DOS 3.31 or higher and connected to an HP LaserJet Series II laser printer.**

**The laser printer should have at least 4 megabytes of extended memory in order to hold the graphics images produced by TRENDS programs.**

## 1.0 PROGRAM EXECUTION:

Programs are activated by typing the program name (without suffix), e.g. to run TRX, simply type TRX and hit the ENTER key. On entry of the desired program name, the computer responds:

TYPE CONTROL FILE NAME:

whereupon the user must enter the name of the file containing the program control information for that particular run, e.g. GRAND.CNT. On hitting the enter key, the program proceeds to execute to completion unless:

1. errors with the control information or the input data set are encountered
2. laser printer problems are encountered.

During the development of the current edition of these programs, numerous checks for illegal control parameters and input data indicators were incorporated into the software. In some cases the programs will stop. Whenever a program crashes, users should immediately check the report file for error messages. Users may encounter program stoppages due to difficulties as yet unencountered by the programs' author.

## 2.0 PROGRAM SIZE LIMITS:

Due to variations in program capabilities, the size limitations on the input data files handled by the respective programs vary as indicated below. The limits given below are for the most compact and fastest versions of the programs. These programs should handle 90% of water quality data series encountered in practice.

TR1: maximum series length = 1000 elements

TRX: maximum series length = 900 elements

TR2: maximum series length = 800 elements

### Very Large Data Sets

Special huge versions of the programs have been compiled to handle very large water quality time series. The huge versions of these programs were achieved by sacrificing error diagnostics, some statistical routines and slower execution times. All data input and program control requirements are the same as for the TR* series. These programs have not been extensively tested.

TH1: maximum series length = 3000 elements

THX: maximum series length = 2500 elements

TH2: maximum series length = 2000 elements

## 3.0 INPUT DATA REQUIREMENTS:

The input data file formats are identical for all 3 programs. Input data must be in the form of an ASCII file formatted according to standard FORTRAN conventions as per the example given in Table 3.

### 3.1 Line 1: Parameter & Site Identifiers

The first line of the input data file should contain:

- an abbreviated water quality parameter name in the first 6 columns, e.g. 'As' (Arsenic) in the Table 3 example or 'COND' (conductance) as in the Table 4 example.
- a 12 character site identification code in columns 7–18, e.g. '170026013 ' in the Table 3 example or 'Clearwater L' as in the Table 4 example.

If either the parameter abbreviation is less that 6 characters or the site identifier is less than 12 characters, these should be left justified to the first allotted column, i.e. columns 1 and 7 respectively.

**N.B.** The parameter name and site identifier in line 1 of the input data file appear on the graphics. This feature was introduced after observing that staff hastily generating graphical results only with test versions of the programs were forgetting to change plot headers and $y$–axis titles and consequently confusing sites and variables. Although not entirely fool proof, this feature provides a checking mechanism in the event that program users change input data files without revising plot header and $y$–axis information.

### 3.2 Lines 2–end of file: Input Data Series

The TRENDS programs' input routines seek the following three variables:

1. sample date read as a 6 digit integer indicating year/month/day, e.g. 641101 for 1964 November 1,
2. sample time read as a 4 digit integer indicating hour/minutes, e.g. 1430 for 2:30 p.m.
3. sample concentration read in FORTRAN real fixed format, i.e. F$x.y$ where $x$ is the total width of the sample concentration field and $y < x$ is the number of digits to the right of the decimal.

In the example data file of Table 3, the first 9 columns of lines $\geq 2$ happens to be occupied by the site identification code. The sample date appears in columns 11–16, the sample time in columns 18–21 and the arsenic concentrations (in $\mu$g/L) appear in columns 22–28. Either of the following FORTRAN formats would lead to data being correctly read by the trend programs:

- (10X,I6,I5,F7.0)
- (10X,I6,1X,I4,F7.0)

where the '10X' skips the reader over the 1st 10 columns of the line and 'I$z$' indicates an integer field of length $z$.

Sample times are of minor importance with sparse, e.g. monthly, sampling; however, when records contain instances of more than 1 sample per day, sample times are important for distinguishing between samples obtained on the same day. In SIS, missing sample times are given the value zero. The trend software converts zero times to 12:00 noon since virtually no PWQMN samples have ever been collected at midnight.

**N.B.** If data records exist for which there are no sample times available, the input file should be prepared with at least 1 column of blanks separating the sample date and the sample concentration. If there were 1 blank column between date and concentration, the input format could be given as (I6,I1,F$x.y$) by which sample times would be read as an I1 column of blanks that are interpreted as zeros and converted to 12:00 noon.

For convenience of numerical manipulation and plotting, the software internally determines a real sample decimal date, e.g. $YR.xxxxx$ where the $YR$ is the year, .e.g. 75, and $.xxxxx$ is the hours, minutes, day and month of the sample expressed as a decimal fraction of the year. Leap years are identified and handled correctly.

**N.B.** When entering input formats in the control (.CNT) files, the parentheses must be included.

**N.B.** The trend programs were designed to work with concentration data having no more than 3 digits to the right of the decimal. The remark applies specifically to cases like Cadmium which if retrieved with SIS default options gives typical sample concentrations as .000$x$ mg/L or .00$xy$ mg/L. Data of this type should always be converted to more appropriate units as $\mu$g/L or ng/L. For PWQMN river data, this can always be done by specifying appropriate unit conversions when retrieving data from SIS. If a data file has been

prepared with concentration data have $\geq 4$ digits to the right of the decimal, a unit conversion feature exists in the .CNT files that would allow such input data to be converted to units with $\leq 3$ digits to the right of the decimal (see PART II, 4.2 Line 4). Note that for data like those presented in Table 4, the trailing digits to the right of the decimal have no effect on the analysis.

**N.B.** Current versions of the trend programs cannot read the ubiquitous alphanumeric remark codes that accompany numerical values of data retrieved from SIS. The '<' symbol appearing to the right of the concentration value of the sample obtained on date 671002 of Table 3 is an example. **If input data format is incorrectly specified and alphanumerics are encountered, the program will crash and return the FORTRAN error message:**

- run-time error F6103: READ(*inputfilename*) – invalid real

### 3.3 Data Matrices:

The input format requirements make it possible to prepare a multi-variable input file for a single site as shown in the example of Table 4 where the date occupies columns 2–7, the time columns 9–12 and the 5 variables were formatted as 5F12.4. The 5 variables happen to be:

1. conductance ($\mu$S/cm @ 25C)
2. hydrogen ion concentration ($\mu$eq/L)
3. calcium concentration ($\mu$eq/L)
4. magnesium concentration ($\mu$eq/L)
5. sodium concentration ($\mu$eq/L)

**N.B.** The missing value code -999.0000 (see PART II: 4.2 Line 6) has been entered in the calcium and magnesium columns as data records for these variables do not begin until 1977.

**N.B.** Line 1 of a multi-variable file is identical to that of a single variable file. When working with an input data matrix, it is more convenient to enter some generic variable indicator in columns 1–6 of the first line as in Table 4, rather than repeatedly enter and edit the variable name, rather then

## 4.0 PROGRAM CONTROL FILES:

Programs execution features are controlled by .CNT (or any suffix the user may prefer) control files. For consistency and ease of use, control files for the three main trend analysis programs are structured similarly, differing only by one or two lines according to the particular requirements of the program. Sample .CNT files for the three programs TR1, TR2 and TRX are given in Table 5. The common elements of the three programs are listed below:

### 4.1 Common Lines of TR1, TR2 & TRX Program Control Files:

1. lines 1–7: all three programs
2. line 8: TR1 [1st 2 elements], TR2 & TRX
3. line 9: TR1 & TR2
4. TRX (line 9) & TR2 (line 10)
5. TR1 & TRX (lines 10–17); TR2 (lines 11-18)

### 4.2 Control File Input Requirements:

**Line 1: all 3 programs; Report File Name**

- 48 character report file name, left justified to column 1

- the report file is read 1st because it is opened immediately to record subsequent information in the .CNT file

**Line 2: all 3 programs; Input Data File Name**

- 48 character input data file name, left justified to column 1

**Line 3: all 3 programs; Water Quality Variable Name**

- 6 character water quality parameter name, left justified to column 1

**Line 4: all 3 programs; Input Data Format**

- 40 character (including parentheses) input data format, left justified to column 1

**Line 5: all 3 programs; Data Bracketing Dates**

- data series bracket dates, i.e. only samples occurring within brackets are read as input
  - lower bracket date as 6 digit integer (year/month/day) in columns 2–7
  - upper bracket date as 6 digit integer (year/month/day) in columns 9–14

**Line 6: all 3 programs; Zero Option & Missing Value Codes**

- 1 digit zero option code [ZOC] (column 1)
  - set to 0 to leave 0's as is
  - set to 1 to replace 0's with 1/2 smallest positive value in record
  - set to 2 to replace 0's with smallest positive value in record
  - set to 3 to delete 0's
- missing value code [MVC] (columns 2-10); format F9.3
  - **N.B.** MVC must be a negative value
  - for data sets with legitimate negative values, e.g. Gran alkalinity & residual series, set MVC to a value smaller than the smallest member of series
  - **N.B.** F9.3 format can be over ridden if necessary, e.g. F9.5, and will be read properly. The desired MVC must appear in columns 2–10 and include the decimal.

**Line 7: all 3 programs; Multiplier/Divisor Codes**

- 1 digit multiplier/divisor code [MDC] (column 1); format A1
  - set to 1 for no multiplier/divisor, i.e. to leave data as is
  - set to M to multiply input data
  - set to D to divide input data
  - **N.B.** Codes M & D must be upper case
- multiplier/divisor (columns 2–10); format F9.3
  - if MDC = 1, can be ignored
  - if MDC = M or D, specify numerical value
  - **N.B.** F9.3 format can be over ridden if necessary, e.g. F9.5, and will be read properly. The desired multiplier/divisor must appear in columns 2–10 and include the decimal.

**Line 8: all 3 programs; Log, Run & Iteration Parameters**

- 1 digit $\log_{10}$ transform code (column 3)
  - set to 0 for no $\log_{10}$ transform
  - set to 1 for $\log_{10}$ transform
- # of program runs (see PART II: 5.2), 3 digit integer or less (preferrably), right justified to column 6
  - set to 2 or 3 for trial runs
- # of iterations in seasonal decomposition loop, 3 digit integer or less (preferrably), right justified to column 9
  - 3 iterations are sufficient for most work

**Line 9: TR1 & TR2; Smoothing Filter Parameters**

- 1 digit smoothing filter identifier (column 2)
  - choose from filters 1 to 9 (PART I: Section 5)
- 1 digit sampling interval code (column 4)
  - choose 1 for median sampling interval (most data sets)
  - choose 2 for mean sampling interval (seasonally sampled lake & river records)
- unit smoothing interval; format F6.2 (columns 5-10)
  - typically 1 or 2 years with TR2

**Line 9 (TRX); Line 10 (TR2); Seasonal Smoothing Parameters**

- minimum number of observations required to compute statistics within seasonal smoothing windows
  - set to 11 or higher (see PART I: Section 5.3 & PART II Section 5.4)
- minimum width of seasonal smoothing window
  - set as a fraction of 1 year, e.g. 12 indicates 1/12 year (see PART I: Section 5.3 & PART II Section 5.4)

**Line 10 (TRX & TR1); Line 11 (TR2); Plot Codes**

- plot code: 1 digit character in column 2
  - set to upper case 'P' for laser printer plots
  - anything else fails to produce laser printer plots
- plot density codes: 1 digit character in column 4
  - set to upper case 'S' for single density plots (75 dots per inch; suitable for fast, rough work)
  - set to upper case 'D' for double density plots (150 dots/inch; suitable for most reports)
  - set to upper case 'Q' for quad density plots (300 dots/inch; suitable for high quality reports & journal articles)
- plot type codes: 1 digit character in column 6
  - set to '0' for ordinary trend plots of either untransformed or log transformed variables
  - set to '1' for logarithmic scaled plot (only permissible when log transform code = 1
- plot annotation code: 1 digit character in column 8
  - set to upper case 'S' for plot without right side annotation
  - set to upper case 'L' for plot with right side annotation

**Line 11 (TRX & TR1); Line 12 (TR2): Plot Header 1**

- up to 56 characters, left justified to column 1
  - appears below main plot header

**Line 12 (TRX & TR1); Line 13 (TR2): Plot Header 2**

- up to 56 characters, left justified to column 1
  - appears below plot header 1

**Line 13 (TRX & TR1); Line 14 (TR2): Current Date**

- up to 20 characters, left justified to column 1
  - e.g. May 31 1991
  - appears in upper right corner

**Line 14 (TRX & TR1); Line 15 (TR2): Y-axis Title**

- up to 30 characters, left justified to column 1
  - if log transform is specified, program automatically supplies '$Log_{10}$' prefix where necessary

**Line 15 (TRX & TR1); Line 16 (TR2): Report Code**

- 2 digit character code left justified to column 1
- specify 'SR' for short report
  - short reports do not include input data list and skip correlation calculations on the final results
  - short reports are recommended
- specify 'LR' for long report
  - long reports include input data list and correlation calculations on the final results
  - the correlations are experimental and may be extremely time consuming
  - long reports are not recommended

**Line 16 (TRX & TR1); Line 17 (TR2): Output File Code**

- 1 digit code left justified to column 1
  - set to '0' for no output files
  - set to any non-'0' 1 digit character for output files

**Line 17 (TRX & TR1); Line 18 (TR2): Trend Output File Name**

- see Section 6.6

**Line 18 (TRX); Line 19 (TR2): Seasonal Output File Name**

- see Section 6.6

### 5.0 TRENDS Primary Program Control Parameters:

There are 3 fundamental program control options that govern the execution of TRENDS programs:

1. log code
2. number of runs
3. number of iterations (not necessary for program TR1).

These are entered in **Line 8** of the program control files of all three programs. These are discussed separately because of their primary importance. The remaining control parameters are discussed in PART II: Section 6.

### 5.1 Data Transformation:

The present TRENDS software permits only two transforms:

- $Z_i = X_i$ (null)
- $Z_i = \log_{10} X_i$

More precise transformations other than logarithmic from the broad Box–Cox family of power transforms might be necessary in certain cases when more formal modelling exercises are undertaken. Because the statistical methods employed in TRENDS programs are highly robust, even if neither the null nor the $\log_{10}$ transforms are the most appropriate, at least one of one of the two available choices should lead to satisfactory graphical results.

Many variables, in particular those exhibiting positive correlation with flow, are positively skewed, i.e. many observations are close to the lower bound and relatively few lie one or more orders of magnitude above the lower values. For positively skewed data, logarithmic transforms commonly lead to data distributions that more closely approximate the normal. Even if raw data are not necessarily positively skewed, when data range over more than 1 order of magnitude, a logarithmic transformation may be advisable as means of obtaining good graphical results. In time series analysis, it is also desirable that local variability exhibited by the data series be relatively constant over the time horizon of the series. Again a log transform may help achieve this objective. However, this rule may not necessarily be apropos if the wide variation in the data series is attributable to:

- only a few outliers

- significant trend in the presence of limited relatively constant local variation.

For the case where wide variability is caused by only a few outliers, it is preferable to leave data as is and rely on the runs feature described below to eliminate the troublesome extreme values. Marginal cases do arise; however, experience suggests that the results of either a null or a $\log_{10}$ transform should be comparable.

When the $\log_{10}$ transform is invoked, it should be kept in mind that models (1–3) are linear additive on the logarithms and that care must be taken with the antilog transformation and the interpretation of the results. With the seasonal adjustment model (3), the relative seasonal component $S_i$ must be added to some trend level, e.g. the time-weighted mean, before antilogs are taken, in order to have meaning on the natural scale. The seasonal amplitude is constant only on the logarithms. On the natural scale seasonal periodicity varies according to whatever level $\tilde{T}$ on which the relative seasonal function has been superimposed. For the written report file, when data have been $\log_{10}$ transformed, the programs TRX and TR2 determine the seasonal amplitude at three trend levels:

1. minimum annual level
2. overall geometric mean trend level
3. maximum annual level

For program TRX, the minimum and maximum annual levels are taken as the years with the smallest and largest medians as determined on the de-seasonalized data, while in program TR2, the respective levels are determined as the yearly time-weighted means of trend estimates $T_i$ occurring within the given years. For TRX, the overall geometric mean trend is the mean of previously cited yearly medians. For TR2, the overall geometric mean level is determined from the grand time-weighted mean trend, calculated with trend estimates $T_i$ as determined for the $\log_{10}$ transformed series. Finally, the results are antilogged to permit users to examine the range of seasonal variation at the three trend levels back in the natural scale.

### 5.2 Runs: Eliminating Extreme Values:

As the final step of a complete analysis or 'run' for a given data set, the three programs TRX, TR1 & TR2 perform an analysis of the model residuals, e.g.

$$\varepsilon_i = X_i - T_i, \qquad \text{(TR1)}$$

$$\varepsilon_i = X_i - T_i - S_i, \qquad \text{(TRX, TR2)}$$

in order to identify extreme outliers which are classified according to various outlier identification criteria. At the final step of the residual analysis, extreme data identified as '**Far Outliers**' or '**Far Out Points**' according to Tukey's (1977) criterion are deleted from the data set. For a normal distribution, Tukey's '**Outer Fences**' coincide with $Q_{50} \pm 4.72\sigma$, where $Q_{50}$ is the median and $\sigma$ is the standard deviation. Of several available choices, this outlier identification criterion is the most conservative in the sense that it is the widest and leads to the deletion of the fewest extrema. Indeed Tukey Outer Fences bound the central 99.99976% of the normal distribution.

The TRENDS programs maintain a running count of completed 'runs' and after a run is complete, the count is checked against a user specified run limit parameter in the main program control line of the .CNT file. If the completed runs are less than the user specified run limit, the analysis is repeated on the reduced data set free of the Far Outliers identified at the end of the previous run. If no far outliers were identified on the previous run, the program stops.

In general there are no hard rules for specifying the number of runs. A minimum of 3 is recommended; however, it is the author's preference to specify a high limit, e.g. 10, and let the programs run to completion. For reporting purposes, analyses may be repeated to some satisfactory level, e.g 2 or 3. Frequently the first round of outlier deletion is essential to eliminate extreme distortion of the graphics $y$-axis scales. Users should not be unduly concerned with the number of Far Outliers identified if these constitute $< 5\%$ of the members of the original data set. If analyses are being performed on untransformed data and the number of Far Outliers approaches 10%, the analysis should be repeated on $\log_{10}$ transformed data.

**It should not be assumed a priori that outliers are erroneous data. They may represent naturally occurring extreme values, e.g. storm runoff concentrations.**

Beyond identifying extreme values for closer scrutiny, outlier elimination on successive runs is intended to improve:

1. graphics scaling
2. non-robust statistics, particularly the estimates of percentage variation contributed by model components.

These features are well illustrated in the example discussed below.

### 5.3 EXAMPLES: $Log_{10}$ Transformation and Runs

An analysis performed with program TR2 of a particular chloride (Cl) data set is illustrated in Figures 11.1–11.4 with no data transformation and in Figures 11.5–11.8 for $\log_{10}$ transformed data. Figures 11.1 and 11.2 show the trend plot and the seasonal plot for the initial Run #1. It is evident in both cases that the overwhelming majority of data are clustered together in a narrow range and the graphical scale has been blown up substantially to meet the requirements of a relatively small number of extreme values. From the trend plot it is evident that virtually all of these have occurred since 1987 and from the seasonal plot it is clear that the extremes tend to occur over the winter, i.e. late in the year and early in the following year.

Figures 11.3 and 11.4 show the comparable plots for Run #2 after the deletion of 16 Far Outliers that were identified during the initial run. The scale is now reduced and the structure in both trend and seasonal data series is clearly evident. The trend plot shows an abrupt rise in Cl levels that appears to begin in 1987. Although rather weak, Cl seasonality conforms with expected behaviour given its usage as winter road de-icer. A review of the 16 outliers in the report file shows the majority of them to have occurred during the winter months.

The other feature to note in comparing Run #1 to Run #2 is the improvement in the estimates of % Variability contributed by trend and seasonal components. On Run #1 trend contributed only about 9% of the variability in seasonally-adjusted residuals while on Run #2 trend contributed about 51% of the variability. Less dramatically on the seasonal plots, the seasonal function contribution to variability in de-trended residuals increases from about 3% to about 14% after deletion of the 16 Far Outliers. The % variability estimates are analogues of the $R^2$ statistic of regression analysis and like $R^2$ are highly sensitive to the presence of extreme values. One of the approaches to developing robust $R^2$ estimates (e.g. Rousseeuw and Leroy, 1987) has been to weight extreme values to 0 before performing the calculations. Zero weighting is virtually equivalent to the deletion mechanism that is built into the TRENDS software.

Figure 11.5–11.8 show the comparable analysis with program TR2 performed on the Cl data series after $\log_{10}$ transformation. The log transform seems justified here as the data clearly span nearly 2 orders of magnitude. Moreover, the results obtained after the initial run are far more informative than those obtained with the untransformed series. Run #2 reveals that only 6 Far Outliers were deleted, another indication that the $\log_{10}$ transform is the better choice. The main improvement of Run #2 occurs with the seasonal function where scatter and consequently the ordinate scale has been reduced by fully 1 order of magnitude.

### 5.4 ITERATIONS: Seasonal–adjustment models (TRX & TR2 only)

The need for an iterative-decomposition scheme to fit seasonal adjustment models arises because the initial guesses at trend term $T_i$ and seasonal term $S_i$ may be rather crude. Furthermore the problems may be aggravated by chronologically or seasonally biased data distributions at various times in the record. The benefits of iterative fitting are illustrated in the following example of a lake total phosphorus (P) series.

#### 5.4.1 ITERATIONS EXAMPLE: Heart Lake Total P, 1975-1988.

This example concerns a nutrient data series gathered in a long term lake eutrophication study. From 1975–1988 samples were collected almost exclusively from May–October during the study years. Almost no samples were collected during the November–April period. Due to the regularly recurrent 5 month gap, the program TRX is considered the most appropriate for analyzing the historical P records. An initial round

of analysis identified 2 extraordinarily high data points that were deleted to produce a simulation of what TRX would produce during the actual Run #2 over successive iterations.

The effects of iterative fitting the trend component are illustrated on the sequence of Figures 12.1–12.5 which correspond respectively to the output after 1, 2, 3, 5 and 10 iterations. Since, on the first iteration the seasonal term is effectively assumed to be zero, Figure 12.1 shows trend determined on the raw data. As we see, from Figure 12.1 to 12.2, there is a dramatic reconfiguration of the trend line on the second iteration where the data have been adjusted by the first guess at the seasonal function. Also note that the estimates of:

- the mean of the annual medians has dropped somewhat from 55.2 to 51.6 $\mu$g/L
- the maximum of the annual medians has dropped significantly from 90 to 77.7 $\mu$g/L
- the estimate of % variability contributed by the trend term to the de-seasonalized residuals has virtually doubled from 10.8 to 21.1%.

In going from Figure 12.2 to 12.3, and finally from Figure 12.3 to 12.4 further adjustment is observed. A final run with 20 iterations produced no noticeable change over the results obtained at 10 iterations which differed only slightly from results obtained with 5 iterations. A review of the report files shows that in 4 years out of 14, the final seasonally adjusted annual median figures after 10 iterations declined by more than 10 $\mu$g/L from the original estimates of annual median based on un-adjusted data.

Figures 12.6, 12.7 and 12.8 show the seasonal plots corresponding to the trend plots of Figure 12.1, 12.2 and 12.5. It is clear that for the P data, the seasonal function on the first iteration (Figure 12.6) gives a reasonably good approximation of the seasonal periodicity. On the second iteration, the main difference is that the seasonal peak has shifted forward by a few days and the data exhibit less scatter about the smoothed seasonal median which accounts for the increase in the % variability explained by the seasonal function. From the second iteration forward, the changes in seasonal function are scarcely perceptible; although, the seasonal amplitude increases slightly to reach its final figure of 40.6 $\mu$g/L at the 5th iteration at which the the % explained variability also stabilizes at about 49.3%.

The changes observed over successive iterations, particularly in the trend estimates, owe their explanation to sampling biases embedded in individual year's data relative to the net seasonal characteristics exhibited by the data. For the 4 years of record for which annual median estimates declined appreciably on iteration, the available observations were biased towards the peak the annual seasonal cycle and were consequently adjusted downwards.

Although the seasonally adjusted annual median medians show appreciable shifting as iterations progress, it should be noted that there remains considerable scatter in the adjusted data; such that, rigorous statistical testing may not show significant year-to-year differences between many of the years which have altered position after the first few iterations.

## 6.0 TRENDS Secondary Program Control Parameters:

### 6.1 Data Series Bracketing:

**Line 5** of all programs .CNT files allows users to enter a pair of 'bracketing dates' which can be used to delimit a segment of the input data series for analysis. Only samples falling between the two dates are accepted for analysis. Reasons for using bracketing dates include:

1. excluding older records of dubious character (see Example 1),
2. excluding stray points far removed from the main series and particularly with program TRX, excluding years at the beginning or end of record with few data points (see Example 2),
3. restricting length of very large data series to meet program size constraints.

Generally, on preliminary runs the bracketing dates should be set to span the full time horizon of the available data records.

### 6.1.1 Bracketing Example 1:

Figures 13.1 and 13.2 illustrate the potential use of bracketing dates. The original bracketing dates were 730000 and 910000. Figure 13.1 shows a Na series for which the first 1 1/2 years of record comprise measurements of dubious character due to limitations of the early laboratory techniques available at the time. A check of the input data file revealed that reasonably reliable Na records begin during July 1974. The final observation of the series occurs in early 1990 and is the only datum for that year. For a data set comprising only a single point, the median defaults to the value of that point. On Figure 13.1, the final annual median for 1990 suggests a continuing upward trend which is not justifiable on the basis of a only a single observation. Figure 13.2 was produced by rerunning the analysis with the bracketing dates 740700 and 900000.

### 6.1.2 Bracketing Example 2:

Figures 14.1 and 14.2 illustrate another example for which series bracketing is used with program TRX. In Figure 14.1, the first datum occurs during 1973, but the record begins in earnest only during 1975. Also at the right end of the Fe series, the final datum occurs during early 1991 and is the only datum available for that year. Hence at both extremities of the Fe series, program TRX determines annual medians for 1973 and 1991 based only on 1 datum in each case. As Figure 14.1 reveals this can create the misleading impression that Fe levels have dropped dramatically in 1990. Figure 14.2 produced by using the bracketing dates 7500000 and 900000 yields a more reliable visual impression.

## 6.2 Zero Option Codes:

**Line 6** of all programs .CNT files permits users to choose from 4 ways of dealing with 0 values that might be present in the data set. Except for a few water quality variables, e.g. temperature and Gran alkalinity, 0's should not be present in water quality data sets. Rather than 0, zero measurements should be reported as < some quantitative value, e.g. a detection limit. Nevertheless 0's may sometimes be encountered in older records and data sets prepared by others. If there are only a few 0's, the user's choice amongst the 4 available options may have little effect on the results. Since logarithms cannot be taken of 0's, the programs will stop and issue an error message if the control file indicates the $\text{Log}_{10}$ transform and 0's are encountered.

## 6.3 Missing Value Code:

The missing value code [MVC] of **Line 6** of all programs permits the user to input data from data matrix files where missing observations have been coded with an MVC as discussed in Input Data File Formats.

## 6.4 Multiplier/Divisor Codes:

The Multiplier/Divisor options of **Line 7** permit users to change data to more appropriate units for analysis. The following two common cases arise:

1. awkward data retrieved at SIS default specifications
2. conversion of ion data from mass concentrations to charge equivalents, i.e. $\mu$g/L to $\mu$eq/L

### 6.4.1 Multiplier/Divisor Example 1: Simple unit conversions

Case 1 is occurs most commonly when users neglect to specify units when retrieving data from SIS. Perhaps the most frequent case arises with the variable total phosphorus (P) which SIS provides at default units of mg/L. Except for heavily polluted sites, P levels in rivers and lakes are generally in the $\mu$g/L range. By working in mg/L units the analyst must carry additional 0's, a decimal point, and if taking logarithms, negative values all of which contribute visual clutter on the $y$–axis of graphics as illustrated in Figure 15.1 (on the Figure 15.1 $y$–axis, the particular scaling factor '$*10^{-1}$' is an uncontrollable foible of the PLOT88 plotting software). By employing the Multiplier/Divisor options to convert the input P series to $\mu$g/L units, Figure 15.2 has much cleaner $y$–axis labelling and annotation.

Figures 15.3 and 15.4 illustrates similar effects of unit selection on certain logarithmic plots. When Fe is expressed in mg/L, the $y$–axis tic mark labels are forced to carry the'–' sign as opposed to the more natural expression in $\mu$g/L on Figure 15.4.

### 6.4.2 Multiplier/Divisor Example 2: Conversion to charge equivalents

Case 2 probably would arise most frequently when dealing with the dilute waters of the Canadian Shield where knowledge of fundamental ionic strength indicators and the principle of electroneutrality are frequently of critical importance. Since most MOE ion data are reported in mass concentration units and the SIS retrieval options do not allow for conversion to charge equivalents, the Multiplier/Divisor options can be employed to convert units. Figure 16.1 was generated using the sulphate ($SO_4^{2-}$) data as supplied in mass concentration units. Developing the results in charge equivalent units may be done as follows. First obtain the equivalent weight of $SO_4^{2-} = 96.0576/2 = 48.0288$. Supposing that we wish to obtain the results in $\mu$eq/L from mg/L, we get the number of $\mu$eq/L per mg/L as $1000/48.0288 = 20.82084$. Then the Multiplier option is invoked by placing 'M' in column 1 of **Line 7** of the .CNT file and multiplier 20.82084 in columns 2–10. Equivalently we could proceed by invoking the Divisor option by placing 'D' in column 1 and the divisor .0480288 in columns 2–10. The results are shown in Figure 16.2 where the '$*10^1$' scaling factor on the $y$-axis is an uncontrollable feature of the PLOT88 plotting software scaling routines. Of course, if a data series were supplied in charge equivalent units and mass concentration units were desired, the Multiplier/Divisor options could be invoked perform the conversion.

## 6.5 Plot Codes:

Plot codes are entered in the .CNT files on **Line 10** programs (TRX & TR1) and **Line 11** program TR2. As indicated earlier, usage of the first plot code 'P' to generate laser printer graphics is straight forward.

### 6.5.1 Plot Density Code

The plot density code entered in column 4 allows users to control the quality of the graphics produced. Usage is illustrated in Figures 17.1–17.3. Figure 17.1 is a **single density** plot obtained by placing '**S**' in column 4. Figure 17.2 is a **double density** plot obtained by placing '**D**' in column 4. Figure 17.3 is a **quad density** plot obtained by placing '**Q**' in column 4. The differences in plot quality are readily evident. It is important to note that the increased quality of higher density plots is rather inversely proportional to the time required to generate the plot. Regardless of the power of the computer used the speed of data transmission to the laser printer remains constant. A quad density plot requires the transmission of approximately 4 times the information of a single density plot; hence, considerable waiting time. To reiterate earlier recommendations:

- Single density plots are appropriate for quick rough work. Numerals and letters may be rather difficult to decipher.
- Double density plots are acceptable for most reporting purposes. Some characters, e.g. '8', are not well formed and may suffer further on reproduction. Almost all graphics in this manual were produced at double density.
- Quad density plots are recommended for high quality reporting and journal article graphics.

### 6.5.2 Plot Type Code

The plot type code entered in column 6 is useful for controlling the type of trend plot generated when the log transform is invoked. When no log transform is requested the plot type code should be set to '0' (an error message is invoked and the program stopped if this occurs). Figures 16–17 are examples of the standard trend plot format for untransformed data. When invoking the log transform, users have the choice of the following 2 trend plot formats:

1. Set to '0' a normal plot is produced with the log transformed data as on Figure 18.1. Note that the $y$–axis is scaled identically as a normal plot.
2. Set to '1' the $y$–axis is scaled in decadal $\log_{10}$ cycles as on Figure 18.2.

The reasons for offering the two log plot types should be evident from a comparison of Figures 18.1 and 18.2. When a given data set extends across two or more decades as in Figure 18.2, it may occur that the

data series is concentrated into the centre 1/3 of the plot leaving about 2/3 of the plot space blank. In contrast, on normally scaled Figure 18.1 the data series fills the available space and provides a superior visual representation of the series behaviour over time. Unfortunately, due to the erratic scaling routines of the PLOT88 software, the normally scaled log plots do not always yield such satisfactory results.

Though some may argue for exclusive usage of decadal style plots as they permit direct reading of numerical values, **illustrating the relative shifts in the data series over time is the primary purpose of these graphics.** The report file provides summaries of annual seasonally adjusted statistics.

Microbiological indicators, e.g. fecal coliforms, are one class of variables which typically exhibit such high variability that the decadal style log plot is almost always most appropriate, as in Figure 18.3 on which the data are seen to extend across 6 decades.

### 6.5.3 Plot Annotation Code

The plot annotation code in column 8 allows the user to retain or delete the textual information and summary statistics appearing on the right side of the plot. An 'L' placed in column indicates a 'long' plot which presents the information on the right side. The primary reason for the right side annotation is to assist users in keeping track of the myriad of graphics that can be generated rather quickly with the software.

An 'S' placed in column indicates a 'short' plot which presents only the plot as in Figure 19. The primary reason for offering this option is to facilitate the cutting and pasting of graphics that is sometimes required when producing reports and articles. Note that the two plot headers and the $y$-axis title may be left blank in the .CNT file as discussed below.

### 6.5.4 Plot Headers, Run Date & Y–axis Title:

The locations of the 2 plot headers (PLOT HEADER 1 & PLOT HEADER 2), the current date (TODAYS DATE) and the $y$–axis title (Y-AXIS TITLE) are illustrated on Figure 20.1 which presents the TR2 analysis of the same microbiological Figure 18.3. Although space is provided for these features, all may be left blank by inserting blank lines at the appropriate locations in the .CNT files. An example is shown in Figure 20.3.

## 6.6 Output Files:

Entering any non–'0' digit in column 1 of **Line 16 (TRX & TR1); Line 17 (TR2)** of the .CNT files generates:

- a 'Trend Output File' for all 3 programs
- a 'Seasonal Output File' for TRX & TR2

according to the names & locations specified in lines 17–18 of the .CNT files. The features of these output files are described below.

### 6.6.1 Trend Output Files:

**Line 1:**

- The first 18 columns of Line 1 of trend output files are identical to those of the input data file, i.e. the parameter identifier in columns 1–6 and the site identifier in columns 7–18.
- To the right of the parameter & site identifiers, beginning in column 23 appear banners identifying the program generating the file, i.e.

  TRX TREND OUTPUT FILE
  TR1 TREND OUTPUT FILE
  TR2 TREND OUTPUT FILE

**Lines 2–end of file: Trend file output**

**N.B.** All trend output files contain the same number of lines the input data file less however many 'far outliers' have been deleted by the final 'run'.

The formats of trend file output are:

- (I7,I5,F12.3,3F12.4) program TR1
- (I7,I5,F12.3,5F12.4) programs TRX & TR2

The 1st three numerical columns of all trend output files contain:

- sample date (yr/mo/dy) in columns 2–7
- sample time (hr/min) in columns 9–12
- sample concentration $X_i$ in columns 13–24 in format F12.3

The 4th column always contains the corresponding trend term $T_i$ or, if the analysis has been performed on $\log_{10}$ transformed data, antilog $T_i = 10^{T_i}$.

The content of the remaining columns to the right vary according to:

- the generating program, i.e. TR1 output differs from that of TRX & TR2
- whether the analysis was performed on natural or log transformed input data

**TR1 OUTPUT TREND FILE: untransformed data**

- output file contains only 4 numerical columns corresponding to the sample date, time, concentration $X_i$ and the trend estimate $T_i$

**TR1 OUTPUT TREND FILE: $\log_{10}$ transformed data**

- the 1st 4 numerical columns contain the sample date, time, concentration $X_i$ and the antilogged trend estimate $10^{T_i}$
- numerical column 5 contains $\log_{10} X_i$
- numerical column 6 contains $T_i$ the trend estimate determined on $\log_{10} X_i$

**TRX & TR2 OUTPUT TREND FILE: untransformed data**

- the 1st 4 numerical columns contains the sample date, time, concentration $X_i$ and the trend estimate $T_i$
- numerical column 5 contains $S_i$ the relative seasonal
- numerical column 6 contains $T_i + S_i$, i.e. the sum of the systematic components estimated by the time series model

**TRX & TR2 OUTPUT TREND FILE: $\log_{10}$ transformed data**

- the 1st 4 numerical columns contain the sample date, time, concentration $X_i$ and the antilogged trend estimate $10^{T_i}$
- numerical column 5 contains $10^{T_i+S_i}$, i.e. the antilogged sum of the systematic components estimated by the time series model
- numerical column 6 contains $\log_{10} X_i$
- numerical column 7 contains $T_i$ the trend estimate determined on $\log_{10} X_i$
- numerical column 8 contains $S_i$ the seasonal estimate determined on $\log_{10} X_i$

### 6.6.2 Trend Output File Use:

The trend output files have several possible uses including:

- detailed examination of the numerical values of $T_i$, $S_i$ and other components for specific samples or time periods
- selective graphics regeneration
- sources files for compilation of special graphics in other software packages, e.g.
  correcting PLOT88 idiosyncratic scaling
  same site, multi-variable overlays

same variable, multi-site overlays

#### 6.6.2.1 Selective Graphics Regeneration:

Presume that the preliminary analysis of a water quality record is allowed to run to completion, i.e. until no further far outliers are deleted, and quick single density graphics are generated. Suppose that completion is achieved on the 4th 'run'. To later regenerate report quality quad density graphics for the 4th run without rerunning all graphics the following procedure may be applied:

1. if there exists a trend output file from the first analysis:
   - specify this as the input data file
   - run the desired program requesting quad density plots
2. if there does not exist a trend output file from the first analysis:
   - rerun analysis specifying 4 runs, no graphics and output files
   - specify the trend output file as the input data file
   - run the desired program requesting quad density plots

These procedures can save considerable time and paper when selected high quality graphics must be regenerated for reports or journal articles.

### 6.6.3 Seasonal Output Files:

The 'seasonal output files' are generated only by programs TRX and TR2.

**Line 1:**

- The first 18 columns of Line 1 of seasonal output files are identical to those of the input data file, i.e. the parameter identifier in columns 1–6 and the site identifier in columns 7–18.
- To the right of the parameter & site identifiers, beginning in column 23 appear banners identifying the program generating the file, i.e.

  TRX SEASONAL OUTPUT FILE
  TR2 SEASONAL OUTPUT FILE

**Lines 2–end of file: Seasonal output file**

The seasonal output file comprise 101 lines with 2 numerical columns containing:

1. the seasonal decimal date series from 0.00 to 1.00 year in steps of .01 year at which the seasonal smooth was determined
2. the smoothed relative seasonal function values for the respective seasonal decimal dates

**The seasonal output file format is: (F7.2,F14.6)**

The seasonal output files are useful for:

- detailed examination of the seasonal function
- source files for:
  - comparative analysis of sites or variables via graphic overlays, generally within other software packages
  - grouping, e.g. cluster analysis, of multi–variable or multi–site seasonal functions

# PART III: REPORT FILES:

## 1.0 Introduction:

All 3 TRENDS programs generate report files that are similar in structure and content, but vary somewhat by program and whether the $\log_{10}$ transform has been invoked. A complete set of 6 report files:

- TRX: untransformed series
- TRX: $\log_{10}$ transformed series
- TR1: untransformed series
- TR1: $\log_{10}$ transformed series
- TR2: untransformed series
- TR2: $\log_{10}$ transformed series

with the analysis of the same nitrate series is available in the Appendix.

**N.B.** In the .CNT files users are given the option of specifying 'short' or 'long' reports. Generally 'short' reports are recommended. The 'long' report includes a listing of input data and correlations amongst model components. Computation of the latter can be extremely time consuming. The correlations are currently under investigation as a tool for diagnosing certain apparently anomalous effects that can be observed in the approximate % contributions to total series variability by model components as discussed in Section 6.2.4.

## 2.0 General Report File Layout:

The content of report files varies according to the time series model, i.e. by program, and whether or not the $\log_{10}$ transformation is invoked. The general layout is as follows:

1. Contents of the program control (.CNT) file are written plus additional information concerning the number of observations. Zero observations are identified and their disposition noted. If invoked, $\log_{10}$ transforms are indicated. The Run number is given. For programs TR1 and TR2, a summary of sampling interval statistics are given and the calculated minimum nearest neighbours to be employed with the smoothing filter.
2. Trend summary
3. Seasonal summary
4. Approximate % contribution to total variation of model components plus summary statistics
5. Model residual summary statistics and outlier identification scan.

When the 'run' parameter is $> 1$ and far outliers have been identified, deleted and the trend analysis repeated, the layout of the following 'run' is identical to items 2–5.

## 3.0 Definitions:

The following definitions are critical to understanding certain of the summary statistics appearing in report files.

**N.B.** Note the following significant notational change from the conventions employed in previous sections:

- the relative seasonal function is now indicated as $S_i'$ in chronologic time or $S_j'$ in seasonal time as defined in Section 7.2.3 below.

### 3.1 Time-Weighted Mean Trend:

For programs TR1 and TR2 employing quasi-continuous trend models, trend summary statistics are developed using **time-weighted mean trends** as a measure of central tendency. For example, the **grand time-weighted mean trend** $\hat{T}$ for the series is determined as

$$\hat{T} = \frac{\sum_{i=1}^{n-1}(T_i + T_{i+1})(t_{i+1} - t_i)}{2(t_n - t_1)}.$$

where $T_i$ is the trend estimate for chronologic time $t_i$. Similar calculations are performed to obtain the **annual time-weighted mean trends** for the year by year trend summaries. The antilogged time-weighted means trends determined on $\log_{10}$ transformed data are referred to as **geometric time-weighted mean trends**.

### 3.2 Relative Trend:

Use of the **relative trend** series, $T_i' = T_i - \tilde{T}$, where $\tilde{T}$ is a grand measure of central tendency in the trend estimates $T_i$ over the time horizon of the series, is convenient for certain algorithmic manipulations and for the determination of the approximate % variation contributed by the trend component. As indicated for the 3 programs, The central tendency measure is taken to be:

- **TRX**: $\tilde{T}$ = mean of annual medians determined on seasonally adjusted residuals $\eta_i = X_i - S_i'$ where $S_i'$ now indicates the relative seasonal
- **TR1**: $\tilde{T}$ = grand time-weighted mean trend $\hat{T}$ determined as per 3.1 on raw or transformed series $X_i$
- **TR2**: $\tilde{T}$ = grand time-weighted mean trend $\hat{T}$ determined as per 3.1 on seasonally adjusted residuals $\eta_i = X_i - S_i'$ where $S_i'$ now indicates the relative seasonal.

### 3.3 Mean Annual Seasonal Median:

In programs TRX and TR2, the relative seasonal component $S_i'$ is not determined exactly as was indicated earlier in PART I: Section 4.2.1, but rather by the following computationally equivalent approach:

1. obtain **de-trended residuals** $\nu_i = X_i - T_i'$ where $T_i'$ is the relative trend
2. reorder de-trended residuals $\nu_i$ by time within the year to obtain the series $\nu_j$
3. apply MSM filter to $\nu_j$ to obtain trial seasonal $\mathcal{S}_j$
4. obtain mean annual seasonal median as $\hat{S} = \sum_{j=1}^{100} \mathcal{S}_j/100$
5. obtain relative seasonal as $S_j' = \mathcal{S}_j - \hat{S}$
6. reorder relative seasonal $S_j'$ to chronologic relative seasonal $S_i'$

This approach yields the mean annual seasonal median estimate $\hat{S}$ which is a grand central tendency measure for the series that is free of seasonal sampling bias. Generally after a few iterations $\hat{S}$ is close to the measure $\hat{T}$ described above. Though convergence is extremely slow, if allowed to continue for many iterations, $\hat{S}$ and $\hat{T}$ gradually become identical for all practical purposes.

### 4.0 Trend Summaries:

### 4.1 Program TRX: untransformed data

As shown on the example in the Appendix, for untransformed data the basic TRX trend summary comprises the 8 numerical columns:

1. year counter
2. year
3. number of observations for that year

4. annual median of seasonally adjusted data, iteration = 1
   - **N.B.** recall from PART I: 4.2.1 & PART II: 5.4, that on iteration 1 the seasonal is assumed to be zero; hence, column 4 contains the annual medians of raw data
5. annual median of seasonally adjusted data, iteration = ITER, the number of iterations specified by the user in the .CNT file
6. the difference between the annual median at iteration = 1 and the annual median at iteration = ITER
   - users can examine the list to determine the effects of iterative decomposition
   - large differences in any year are indicative of significant seasonal sampling bias for that year
7. this column labelled 'YR-TO-YR', simply indicates the year to year difference in final annual medians that appears in column 8
8. the difference between annual medians for the years indicated in column 7
   - provided for users convenience in determining successive year to year jumps of the trend line

Beneath the trend summary are listed:

- the means of the annual medians for iterations 1 and ITER
- for iteration = ITER, the maximum annual median, the minimum annual median and their difference

#### 4.1.2 Spearman's Rho Trend Test

The final component of the trend summary is a simple Spearman's Rho trend test on the final sequence of annual medians obtained on iteration = ITER. The indicated $p$(RHO) value which also appears on the graphics is the significance level for a two sided test. As per the usual conventions, for *ple*.05, Rho is considered to be significant at 95% confidence. This is an extremely crude test that should not be relied on alone. For programs TR1 and TR2, the test results should only be considered when USI = 1 and possibly USI = 2. Future editions of the software will provide additional formal tests.

### 4.2 Program TRX: $\text{Log}_{10}$ transformed data

The TRX summary for $\log_{10}$ transformed data provides a year by year summary of the results as determined in the $\log_{10}$ space followed by a similar summary of the antilogged results. The antilogged mean of the annual medians determined on the logarithms is the **geometric mean annual median.**

### 4.3 Program TR1: untransformed data

The TR1 trend summary comprises:

1. grand time-weighted mean determined over the quasi-continuous trend line
2. a listing of successive local maxima and minima over the trend line followed by the overall extrema and their date of occurrence
3. a year by year summary of annual time-weighted mean trend levels similar to the summary of program TRX
4. Spearman's Rho trend test of the sequence of annual time-weighted mean trends

### 4.4 Program TR1: $\text{Log}_{10}$ transformed data

Similar to untransformed summary, but with important results replicated after antilogging.

### 4.5 Program TR2: untransformed data

Like the TR1 summary the TR2 summary presents:

1. grand time-weighted mean determined over the quasi-continuous trend line
2. a listing of successive local maxima and minima over the trend line followed by the overall extrema and their date of occurrence
3. a year by year summary of annual time-weighted mean trend levels
4. Spearman's Rho trend test of the sequence of annual time-weighted mean trends

### 4.6 Program TR2: $Log_{10}$ transformed data

Similar to untransformed summary, but with important results replicated after antilogging.

### 5.0 Seasonal Summaries:

Seasonal summaries appear only in report files generated by programs TRX and TR2. Since the same procedures are employed, the formats are identical.

### 5.1 Seasonal Summaries: untransformed data

The summary comprises;

1. the mean annual seasonal median as per Section 3.3
2. a scan of local extrema over the seasonal trend line
3. summary of global extrema over the seasonal trend line including the seasonal amplitude determined as the annual maximum − annual minimum

### 5.2 Seasonal Summaries: $Log_{10}$ transformed data

The summary comprises;

1. the mean annual seasonal median as per Section 3.3 in $\log_{10}$ units and backtransformed to natural units
2. a scan of local extrema over the seasonal trend line
3. summaries of global extrema over the seasonal trend line including the seasonal amplitude determined as the annual maximum − annual minimum
   these are determined at three levels:
   a. geometric mean annual seasonal median
   b. maximum annual time-weighted mean trend
   c. minimum annual time-weighted mean trend
   so that analysts may examine how seasonal amplitude changes with trend level with the geometric model, i.e. $\log_{10}$ transformed data

### 6.0 Approximate % Contributions to Variation & Summary Statistics:

The summary reports for all programs contain estimates of the approximate contributions to variation by model components. These estimates reflect variability about grand robust central tendency measures of the series $X_i$, hence approximate. Also a subtractive rather than additive principle is used.

### 6.1 TR1

For program TR1, approximate % contributions to total variation in the raw or transformed data series $X_i$ are determined as follows:

- total variation = $\sum_{i=1}^{n}(X_i - \hat{T})^2$
- variation due to trend = $\sum_{i=1}^{n}(X_i - \hat{T})^2 - \sum_{i=1}^{n}(X_i - \hat{T} - T_i')^2$
- variation due to residual noise = total variation − variation due to trend

where the series is considered to be centred about the grand time-weighted mean trend $\hat{T}$. The variation due to trend, when expressed as a percentage of total variation is equivalently an approximate $R^2$ measure for the model.

### 6.2 TRX & TR2

For programs TRX & TR2 approximate contributions to variation are determined for the 3 series:

1. de-seasonalized residuals $\eta_i = X_i - S_i'$

- relative trend component $T_i'$
- noise component $\epsilon_i = \eta_i - \hat{T} - T_i'$

2. de-trended residuals $\nu_i = X_i - T_i'$
   - relative seasonal component $S_i'$
   - noise component $\epsilon_i = \eta_i - \hat{S} - S_i'$
3. raw or transformed series $X_i$
   - relative trend component $T_i'$
   - relative seasonal component $S_i'$
   - relative systematic components $T_i' + S_i'$
   - noise component $\varepsilon_i = X_i - \hat{X} - T_i' - S_i'$

where:

- relative trend $T_i'$ and central tendency measure $\hat{T}$ are determined as in Section 3.2
- relative seasonal $S_i'$ and central tendency measure $\hat{S}$ are determined as in Section 3.3
- central tendency measure $\hat{X}$ is taken as $(\hat{T} + \hat{S})/2$

### 6.2.1 Variation in de-seasonalized series:

Approximate % contributions to total variation in the seasonally adjusted series $\eta_i = X_i - S_i'$ are determined as follows:

- total variation = $\sum_{i=1}^{n}(\eta_i - \hat{T})^2$
- variation due to trend = $\sum_{i=1}^{n}(\eta_i - \hat{T})^2 - \sum_{i=1}^{n}(\eta_i - \hat{T} - T_i')^2$
- variation due to residual noise = total variation − variation due to trend

### 6.2.2. Variation in de-trended series:

Approximate % contributions to total variation in the seasonally adjusted series $\nu_i = X_i - T_i'$ are determined as follows:

- total variation = $\sum_{i=1}^{n}(\nu_i - \hat{S})^2$
- variation due to trend = $\sum_{i=1}^{n}(\nu_i - \hat{S})^2 - \sum_{i=1}^{n}(\nu_i - \hat{S} - S_i')^2$
- variation due to residual noise = total variation − variation due to seasonal

### 6.2.3 Total variation in raw or transformed series:

Let $\hat{X} = (\hat{T} + \hat{S})/2$ be a grand robust central tendency measure of the series $X_i$ over its time horizon. Approximate % contributions to total variation in the raw or transformed series $X_i$ are determined as follows:

- total variation = $\sum_{i=1}^{n}(X_i - \hat{X})^2$
- variation due to trend = $\sum_{i=1}^{n}(X_i - \hat{X})^2 - \sum_{i=1}^{n}(X_i - \hat{X} - T_i')^2$
- variation due to seasonal = $\sum_{i=1}^{n}(X_i - \hat{X})^2 - \sum_{i=1}^{n}(X_i - \hat{X} - S_i')^2$
- variation due to trend + seasonal = $\sum_{i=1}^{n}(X_i - \hat{X})^2 - \sum_{i=1}^{n}(X_i - \hat{X} - T_i' - S_i')^2$
- variation due to residual noise = total variation − variation due to trend + seasonal

### 6.2.4 Interpreting % Contributions to Series Variation:

The estimated % contributions to series variability are approximate figures that are informative but should nevertheless be treated with caution. Like the $R^2$ statistic of regression analysis, the percentages are non-robust and highly sensitive to the presence of outliers. Thus, if outliers are present, the %'s only begin to

assume a reasonable degree of validity after 2 or 3 'runs' have been performed. Also recall the influence of iterations as in the example of PART II: 5.4.1 where successive iterations significantly increase the % contributions to variability.

For the nitrate series of the sample TRX report files given in the appendix, readers will note that the 'TOTAL VARIATION IN SERIES' summary for untransformed data reveals:

- variation due to trend = 58.3 %
- variation due to seasonal = -5.8 %
- variation due to trend + seasonal = 68.4 %.

Most often the separate trend and seasonal contributions should add to produce a figure approximately equal to the joint contribution. Nevertheless, the above figures are not necessarily in error! Rather what they reveal are certain synergisms and biases amongst components embedded in the series. This type of situation may reflect:

- significant correlation between estimated trend and seasonal components
- local residual variability that varies over time, e.g. the case of older records that show much more scatter than current records
- local residual variability that varies with the magnitude of the underlying process, e.g. variability is positively correlated to trend level
- chronologically varying seasonal sampling biases, e.g. older records reflect intensive summer sampling while recent records are evenly distributed throughout the year when a strong seasonality is present

and other causes. To degrees such circumstances represent violations of the general assumptions that underlie the time series models; although, since the models are inherently robust, results concerning trends do not suffer. The correlation coefficients that are determined when a 'long report' is requested are under development as a tool for diagnosing the causes of such phenomena. This is an area requiring further research.

### 6.3 Summary Statistics:

Just beneath the approximate % contributions to variation appear summary statistics determined on the series under examination and the corresponding residual noise. The mean and standard deviation require no explanation, but the robust statistics may be unfamiliar to some. Consider robustness to imply resistance to distribution shape and the presence of extreme values. Definitions follow:

**median:**

- 2nd quartile $q_2$ or 50th quantile $Q_{50}$; a robust measure of central tendency

**MAD:**

- median of absolute deviations from a robust central tendency measure $\tilde{X}$
- herein let $\tilde{X}$ = median $\{X_i\}$, hence MAD $X$ = median $\{|X_i - \tilde{X}|\}$.
- MAD is a robust measure of dispersion $\approx 2/3\sigma$ the standard deviation when data are normally distributed

**IQR:**

- interquartile range, i.e. $q_3 - q_1$ where $q_3$ = 3rd quartile = $Q_{75}$ = 75th quantile and $q_1$ = 1st quartile = $Q_{25}$ = 25th quantile
- IQR/2 is a robust measure of dispersion $\approx 2/3\sigma$ the standard deviation when data are normally distributed

**Robust Standard Deviation:**

- $\sigma^* = S^*/\Phi^{-1}(.75) = 1.4826022S^*$ where $S^*$ is a robust measure of dispersion such as MAD or 1/2 IQR
- the reported robust standard deviation is derived with MAD

- N.B. $\Phi(\cdot)$ is the normal distribution function, thus $\sigma^*$ assumes that robust dispersion $S^*$ is derived from data that over the central 50% of their distribution behave approximately normally

**N.B.** If all $X_i$ are of the same sign, median $\{X_i\}$ and median $\{|X_i|\}$ will be identical

## 7.0 RESIDUAL SUMMARY STATISTICS & OUTLIER IDENTIFICATION:

Finally for the model residuals $\varepsilon_i$, basic statistics, including order statistics, i.e. quantiles, and range statistics, are summarized and an outlier analysis is performed to categorize extreme residuals by several classifications in common use. Far outliers identified as exceeding Tukey's (1977) outer fences are deleted from series $X_i$ in preparation for a repeat trend analysis on a subsequent 'run'. An outer fence is $7S^*$ where $S^* = IQR/2$ which for normally distributed data is about 4.72 standard deviations from the mean.

## ACKNOWLEDGEMENTS

Aspects of the statistical work were supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) operating grant held by the author as Adjunct with the Department of Statistical and Actuarial Sciences, The University of Western Ontario, London, Canada.

## REFERENCES

**Bodo, B.A.** 1989. Robust graphical methods for diagnosing trend in irregularly spaced water quality time series. *Environ. Monitoring Assessment* **12**:407–428.

**Bodo, B.A.** 1991. Trend analysis and mass-discharge estimation of atrazine in southwestern Ontario Great Lakes tributaries: 1981-89. *Environ. Toxicol. Chem.* to appear August.

**Chambers, J.M., W.S. Cleveland, B. Kleiner** and **P.A. Tukey.** 1983. *Graphical Methods for Data Analysis*, Duxbury, Boston.

**Cleveland, W.S.** 1979. Robust locally weighted regression and smoothing scatterplots, *J. Am. Stat. Assoc.* **74**:829–836.

**Cleveland, W.S.** 1985. *The Elements of Graphing Data*, Wardsworth, CA.

**Hastie, T.J.** and **R.J. Tibshirani.** 1990. *Generalized Additive Models*, Chapman and Hall, New York.

**Hipel, K.W., A.I. McLeod** and **R.R. Weiler.** 1988. Data analysis of water quality time series in Lake Erie. *Water Resour. Bull.* **24**:533-544.

**Hirsch, R.M.** and **J.R. Slack.** 1984. A nonparametric trend test for seasonal data with serial dependence. *Water Resour. Res.* **20**:727-732.

**Hirsch, R.M., R.B. Alexander** and **R.A. Smith.** 1991. Selection of methods for the detection and estimation of trends in water quality. *Water Resour. Res.* **18**:107-121.

**Hirsch, R.M., J.R. Slack** and **R.A. Smith.** 1982. Techniques of trend analysis for monthly water quality data. *Water Resour. Res.* **18**:107-121.

**McLeod, A.I.** and **K.W. Hipel.** 1990. *Trend analysis methodology for water quality time series.* R.A.C. Rep. 278C, Ontario Ministry of the Environment, Toronto.

**McLeod, A.I., K.W. Hipel** and **B.A. Bodo.** 1991. Trend analysis methodology for water quality time series. *Environmetrics* to appear.

**McLeod, A.I., K.W. Hipel** and **F. Camacho.** 1983. Trend assessment of water quality time series. *Water Resour. Bull.* **19**:537-547.

**Rousseeuw, P.J.** and **A.M. Leroy.** 1987. *Robust Regression and Outlier Detection*, Wiley, New York.

**Tufte, E.R.** 1983. *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, Connecticut.

**Tukey, J.W.** 1977. *Exploratory Data Analysis*, Addison-Wesley, Reading, Mass.

**Van Belle, G.** and **J.P. Hughes.** 1984. Nonparametric tests for trend in water quality. *Water Resour. Res.* **20**:127-136.

# APPENDIX A

# FIGURES

for

PC-Software, Users Guide and Documentation

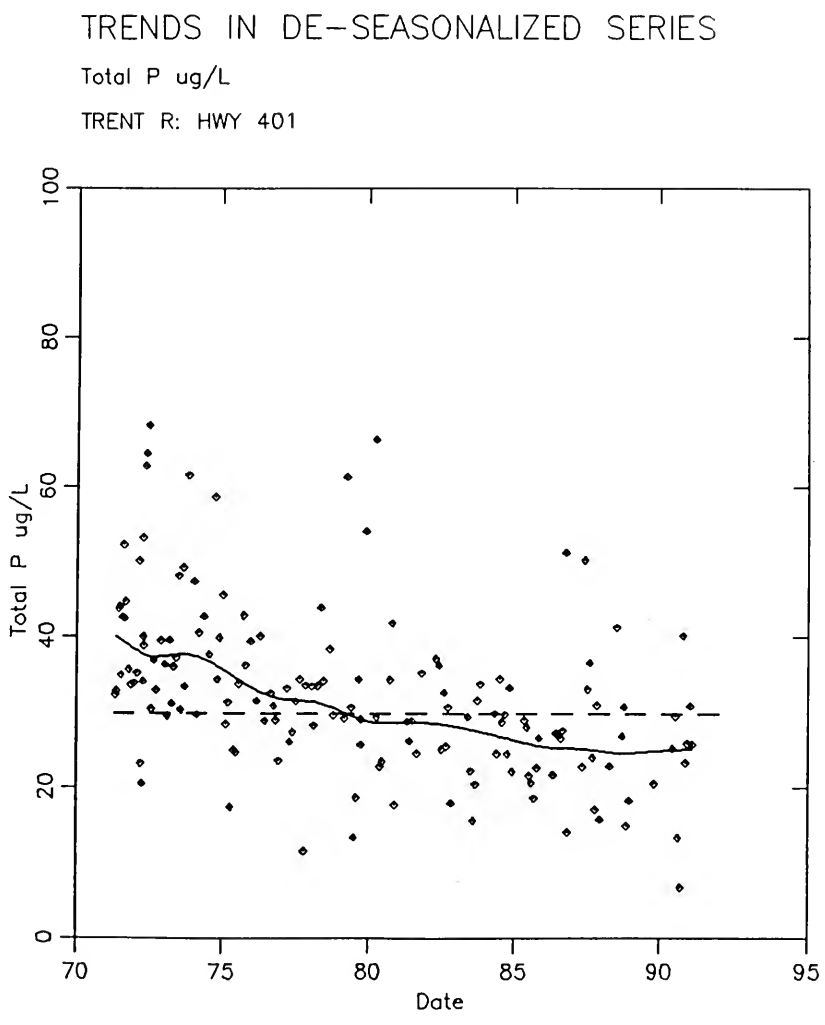
for

Robust Graphical Time Series Analysis

of

Long Term Surface Water Quality Records

**Figure 1a. A smooth trend plot produced by program TR2.**




Run #2: 8 FAR OUTLIERS DELETED
Run date: Jun 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 29.767

Maximum trend = 39.982
@ MAY 12 1971

Minimum trend = 24.530
@ SEP 7 1988

Trend range = 15.452

Observations = 157
Series begins @ MAY 12 1971
Series ends @ FEB 11 1991

Filter = 5
Unit smoothing interval = 2 yr
Minimum smoothing on 25 observations

Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.998
p(RHO) = 0.000 dof = 17

% Variation due to trend = 26.5

Data file: TR45.DAT
Parameter: P
Site: TRENT R HWY

**Figure 1b. A smooth seasonal plot produced by program TR2.**

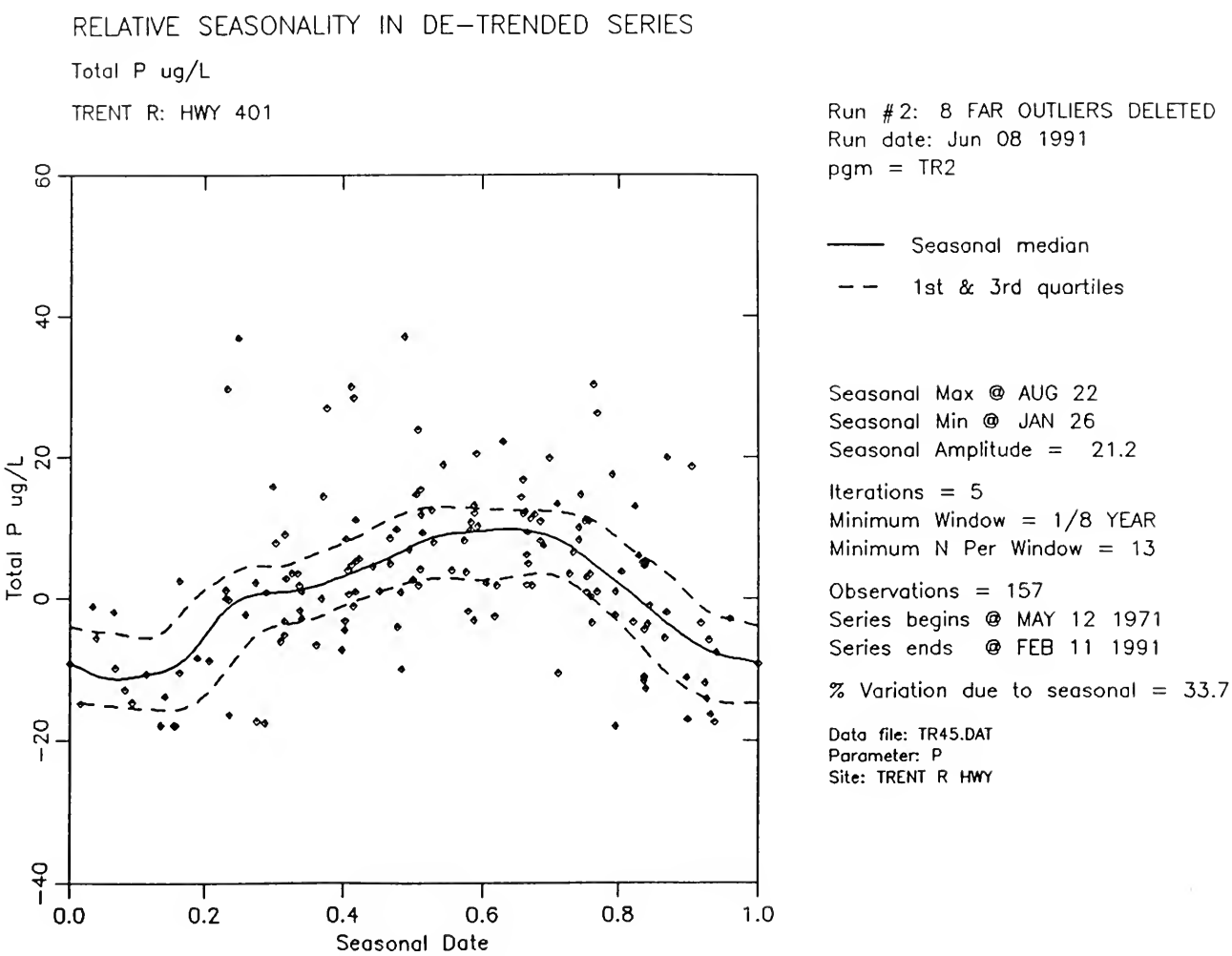


**Figure 2. Trend line estimated as the discrete stepwise annual median filter.**

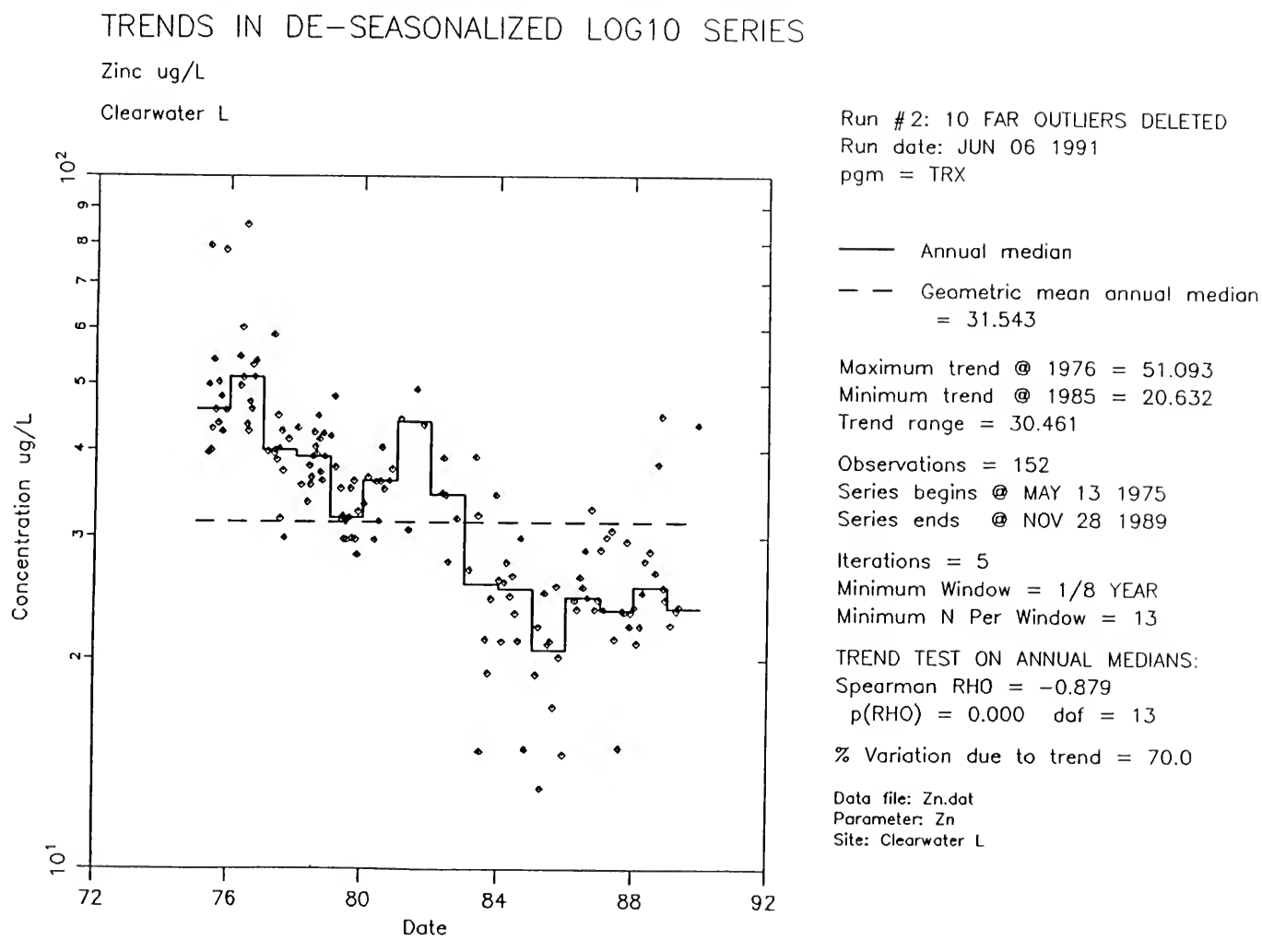


**Figure 3. Trend line estimated with simple Tukey filter.**

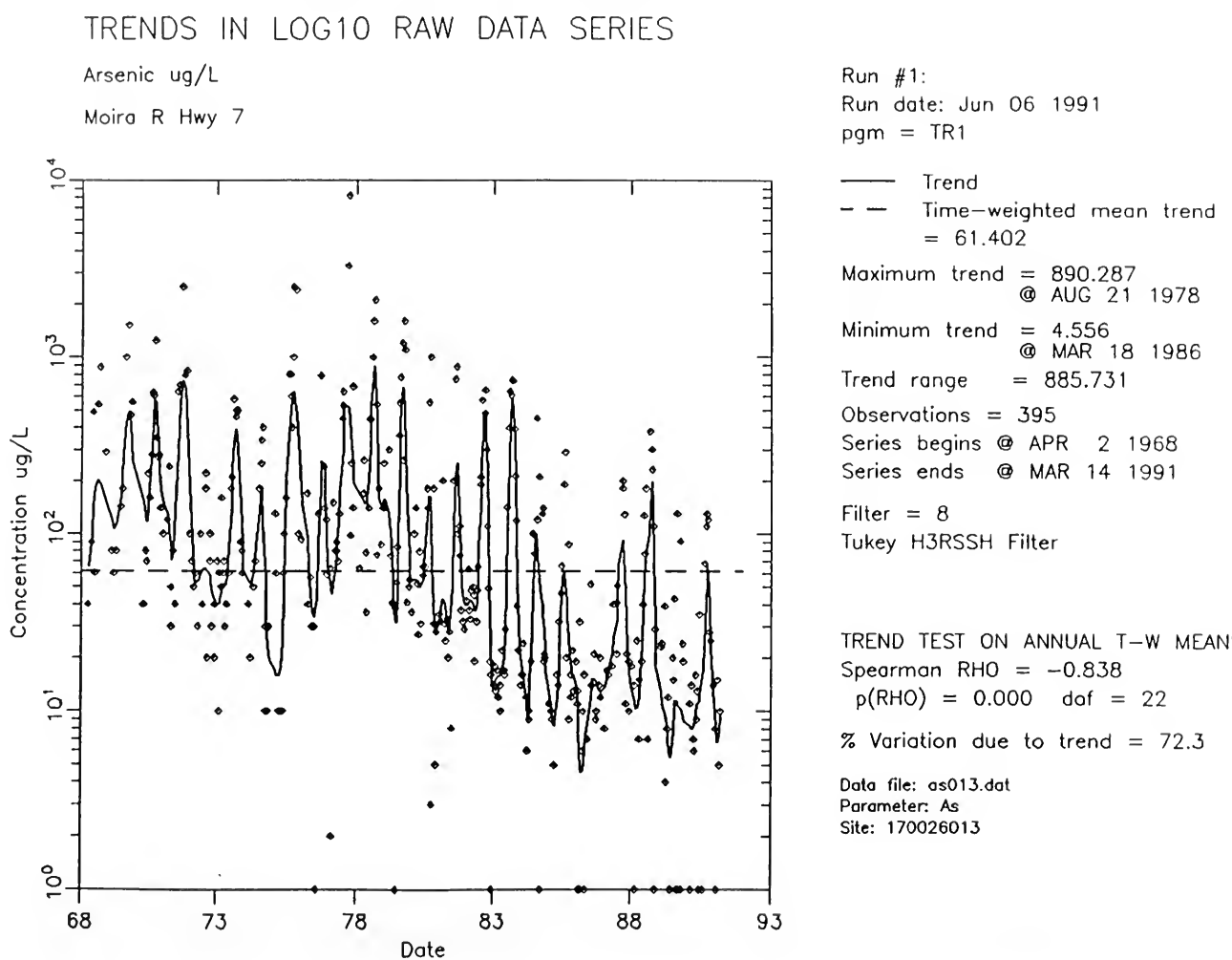


**Figure 4.1 Conductance trends: Filter 1.**

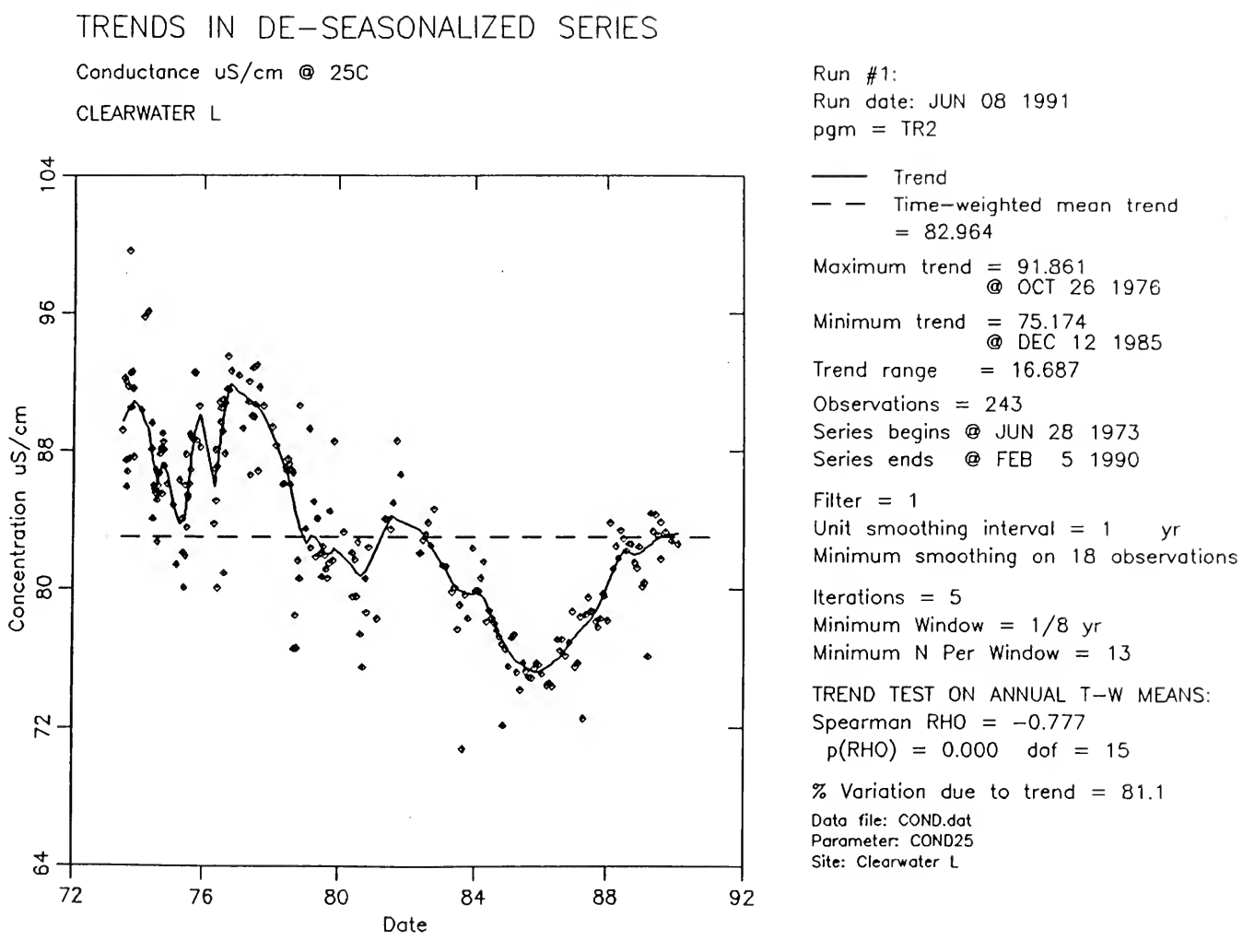


**Figure 4.2 Conductance trends: Filter 2.**

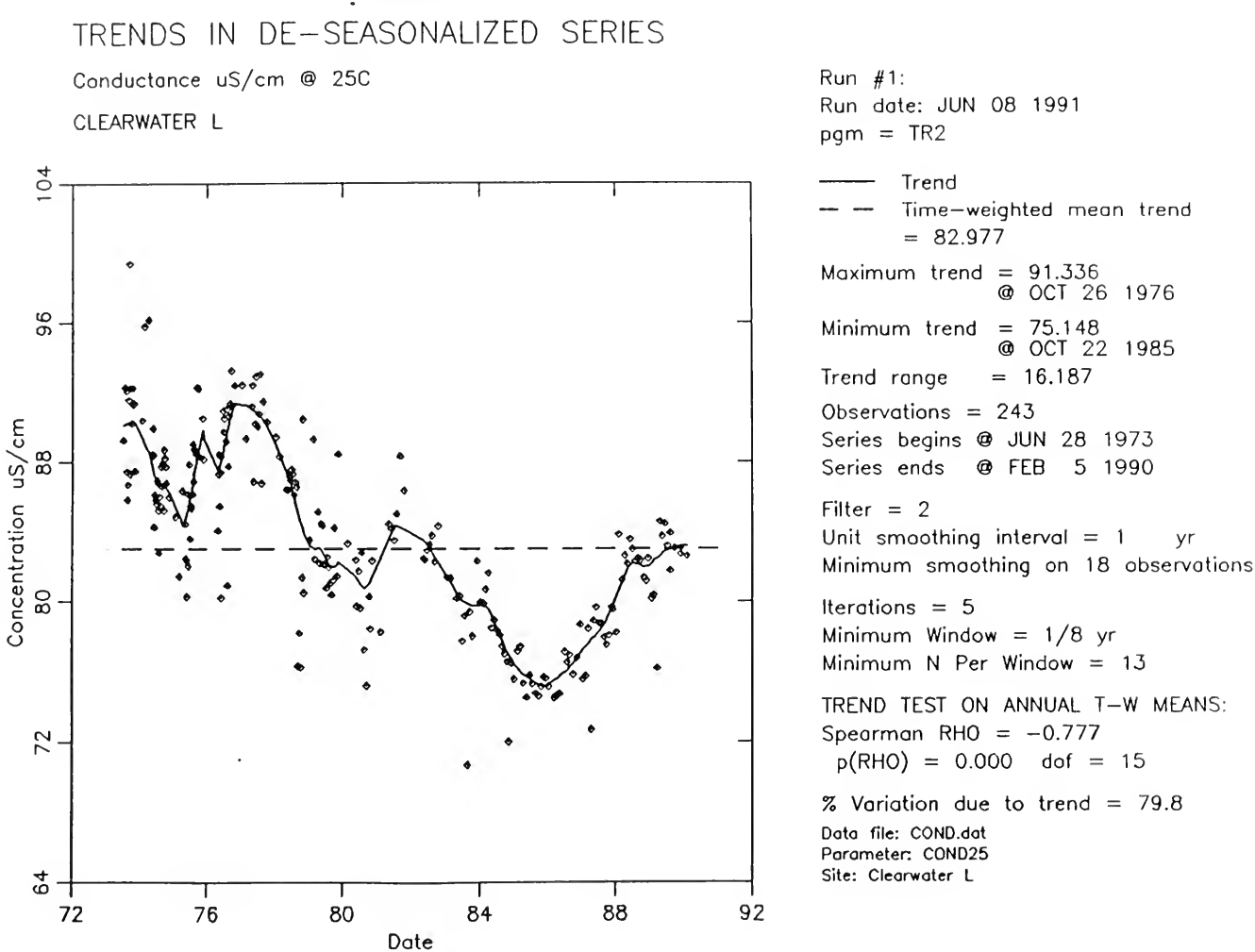
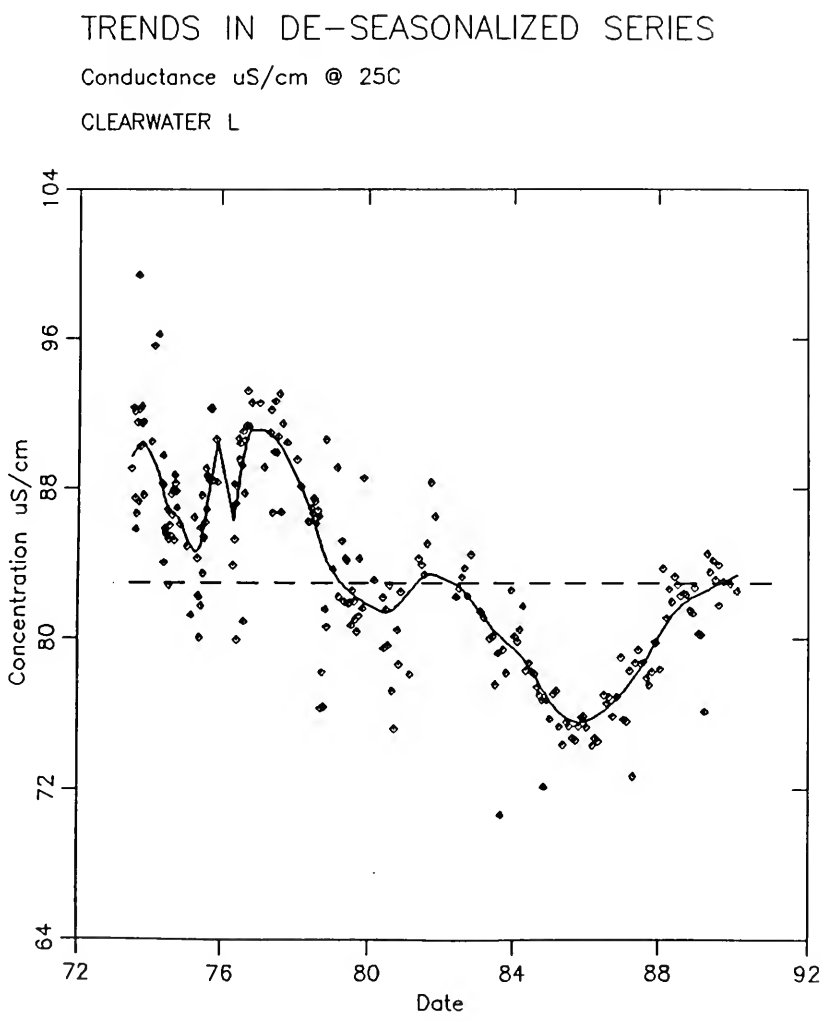


**Figure 4.3 Conductance trends: Filter 3.**




Run #1:
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 82.934

Maximum trend = 91.119
@ JAN 17 1977

Minimum trend = 75.540
@ OCT 22 1985

Trend range = 15.579

Observations = 243
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 3
Unit smoothing interval = 1 yr
Minimum smoothing on 18 observations

Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.779
p(RHO) = 0.000 dof = 15

% Variation due to trend = 79.0

Data file: COND.dat
Parameter: COND25
Site: Clearwater L

**Figure 4.4 Conductance trends: Filter 4.**

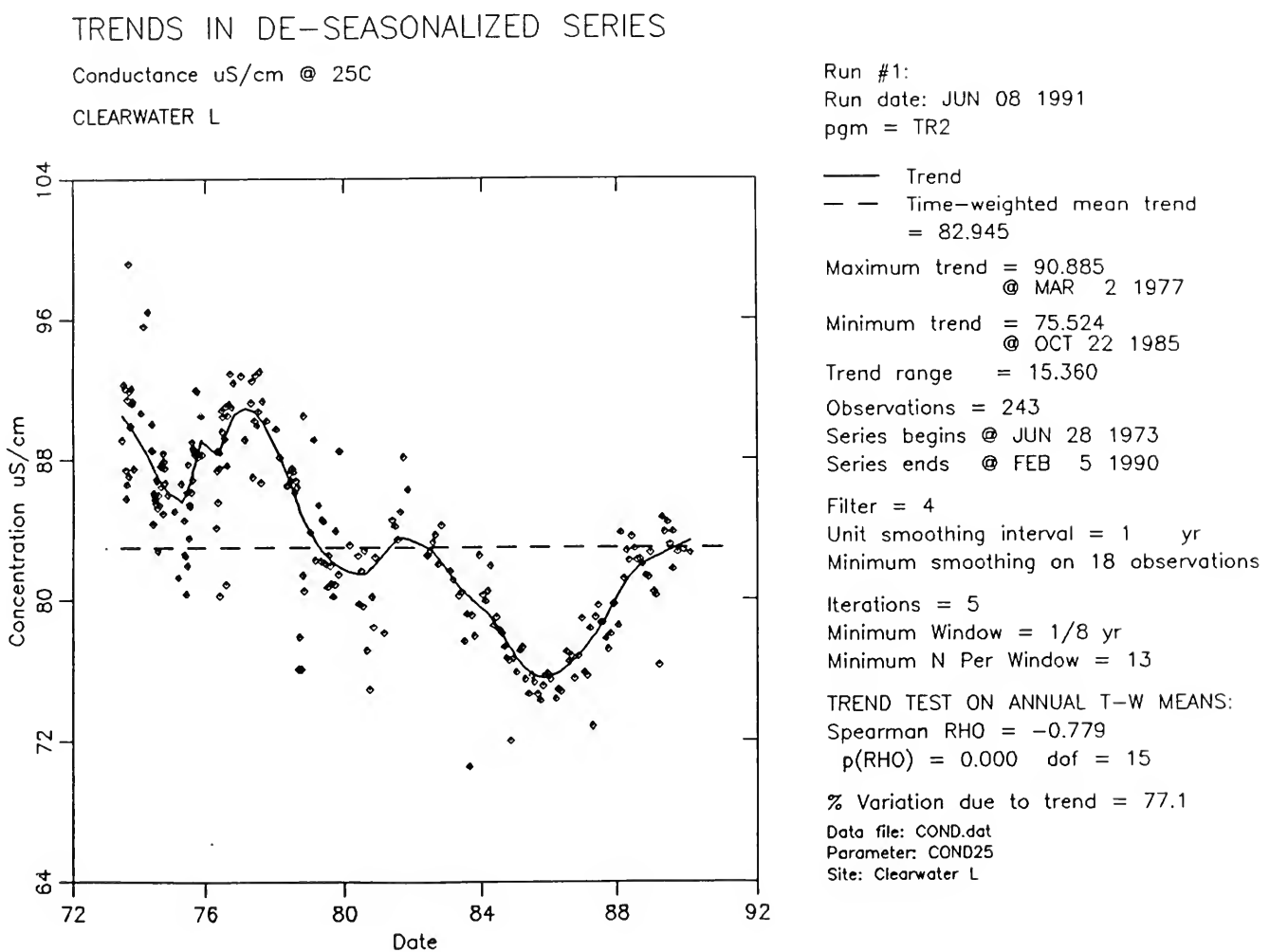


**Figure 4.5 Conductance trends: Filter 5.**

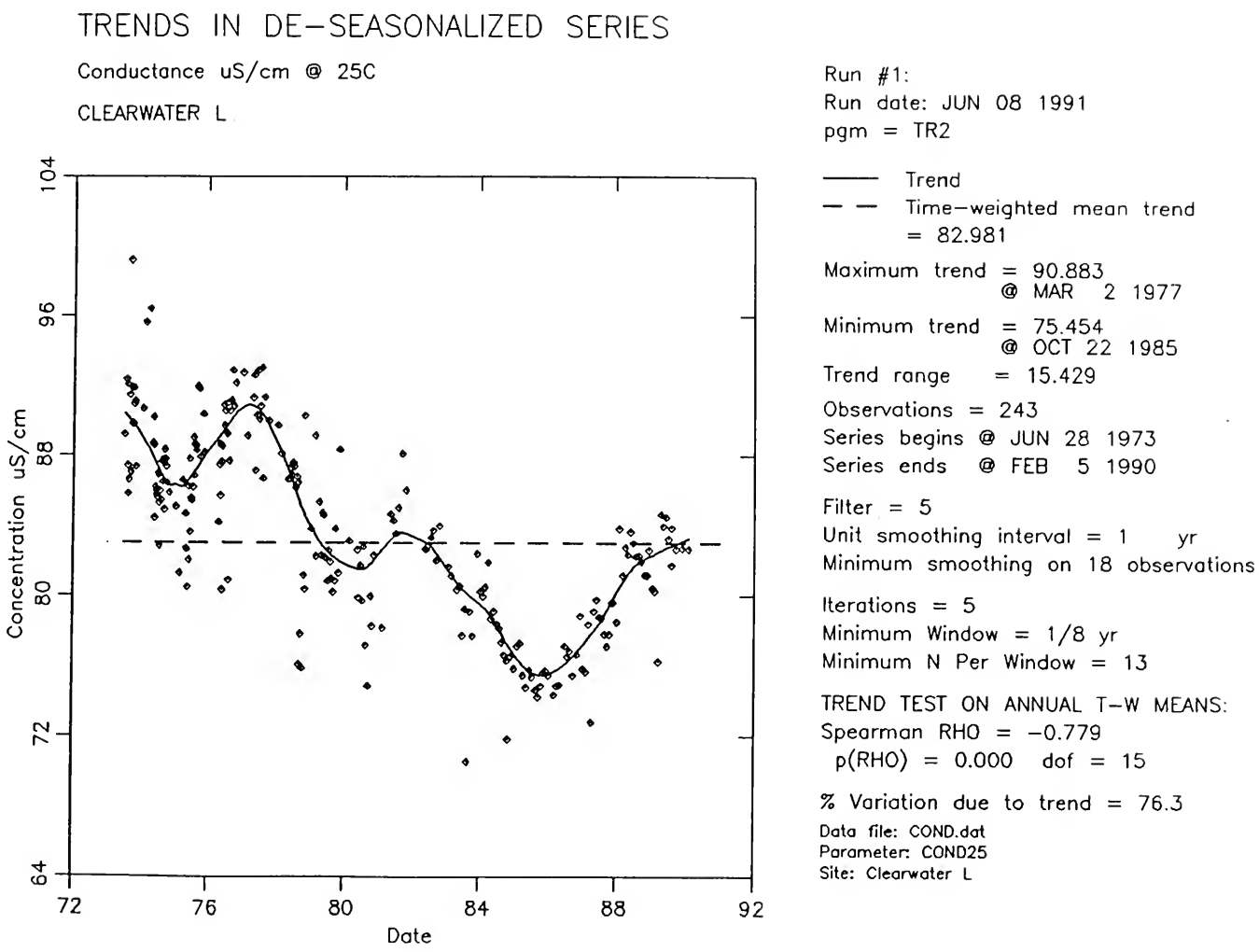


**Figure 4.6 Conductance trends: Filter 6.**

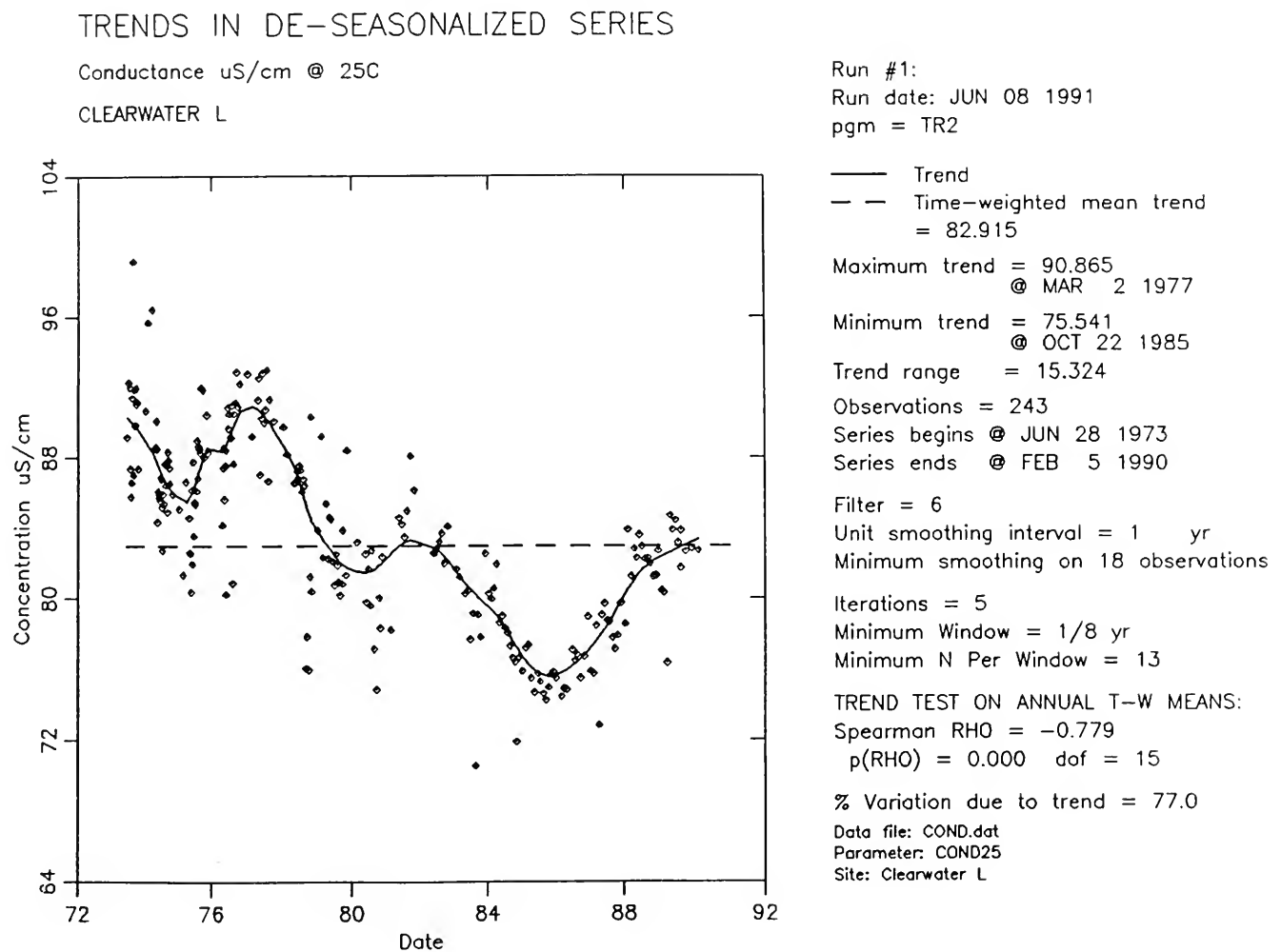


**Figure 4.7 Conductance trends: Filter 7.**

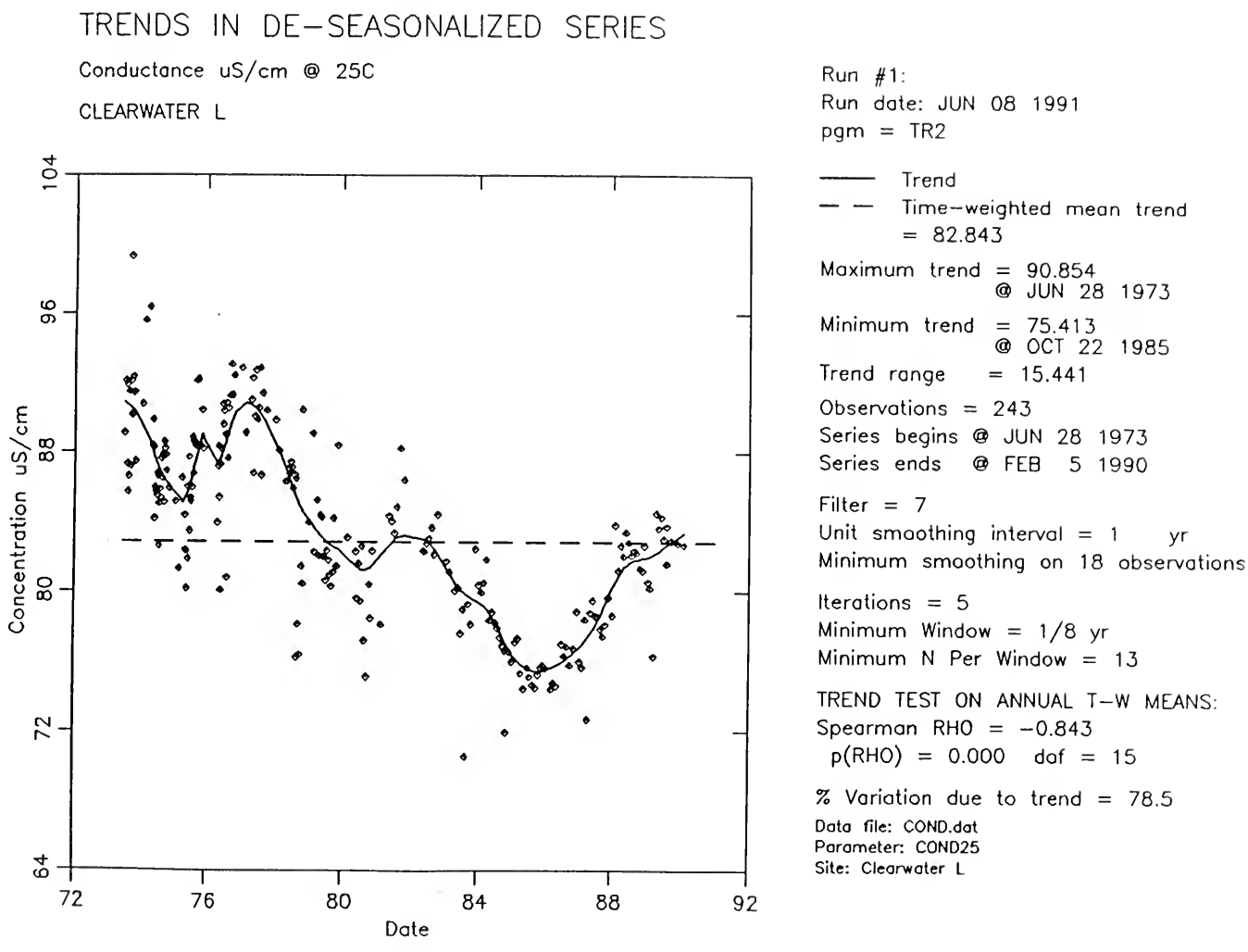


**Figure 4.8 Conductance trends: Filter 8.**

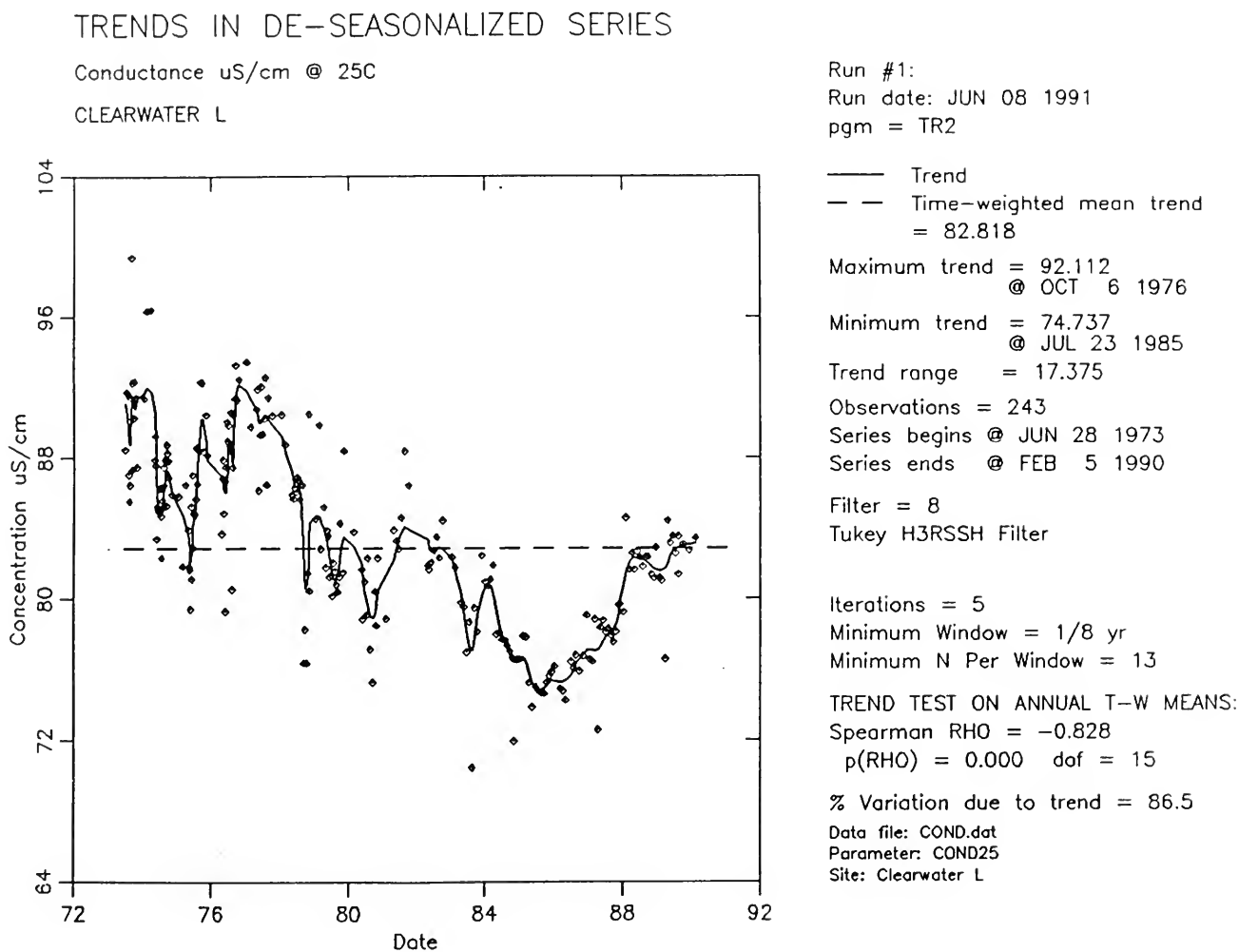


**Figure 4.9 Conductance trends: Filter 9.**

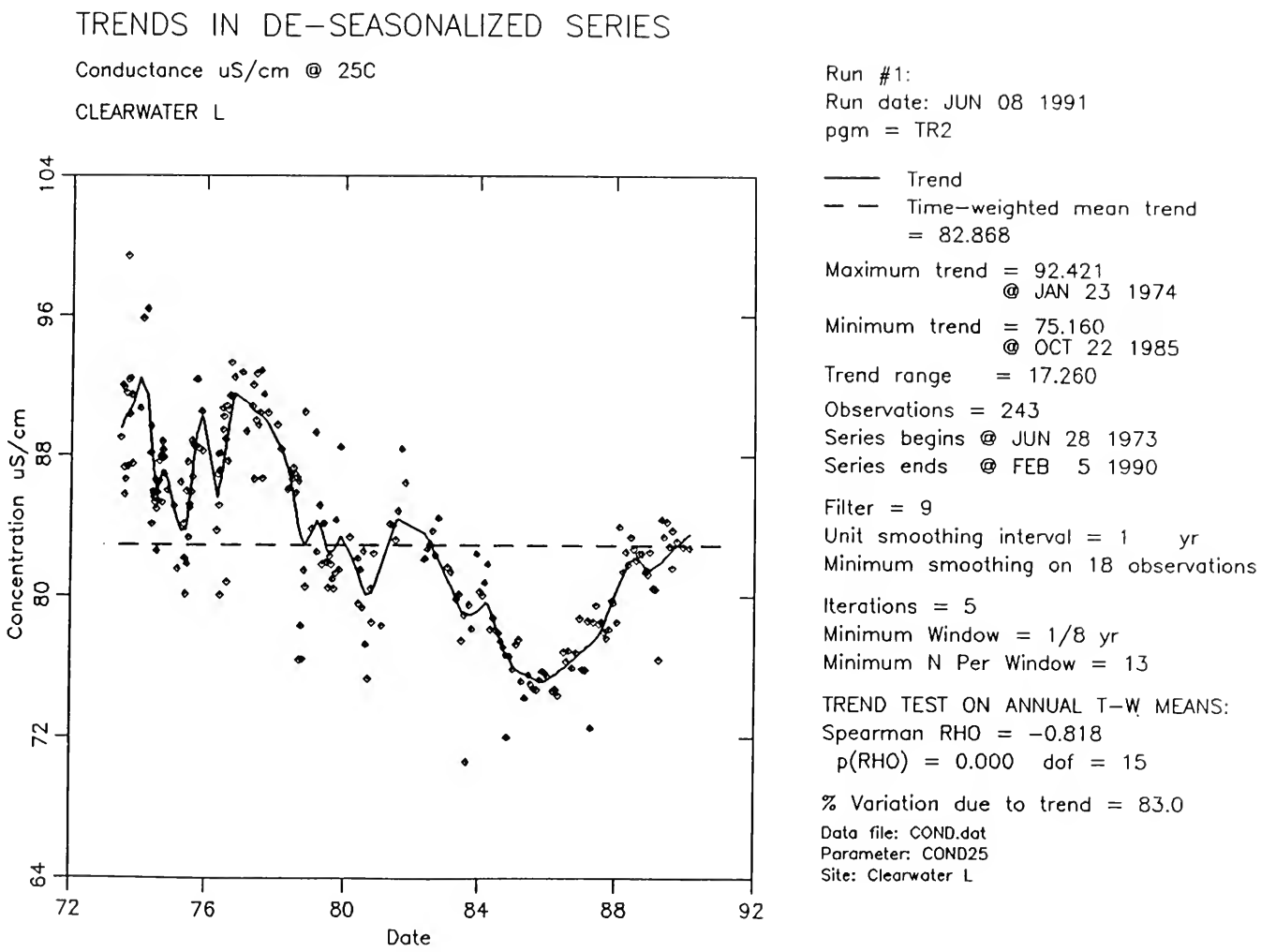


**Figure 4.10 Conductance trends: mean annual median filter.**

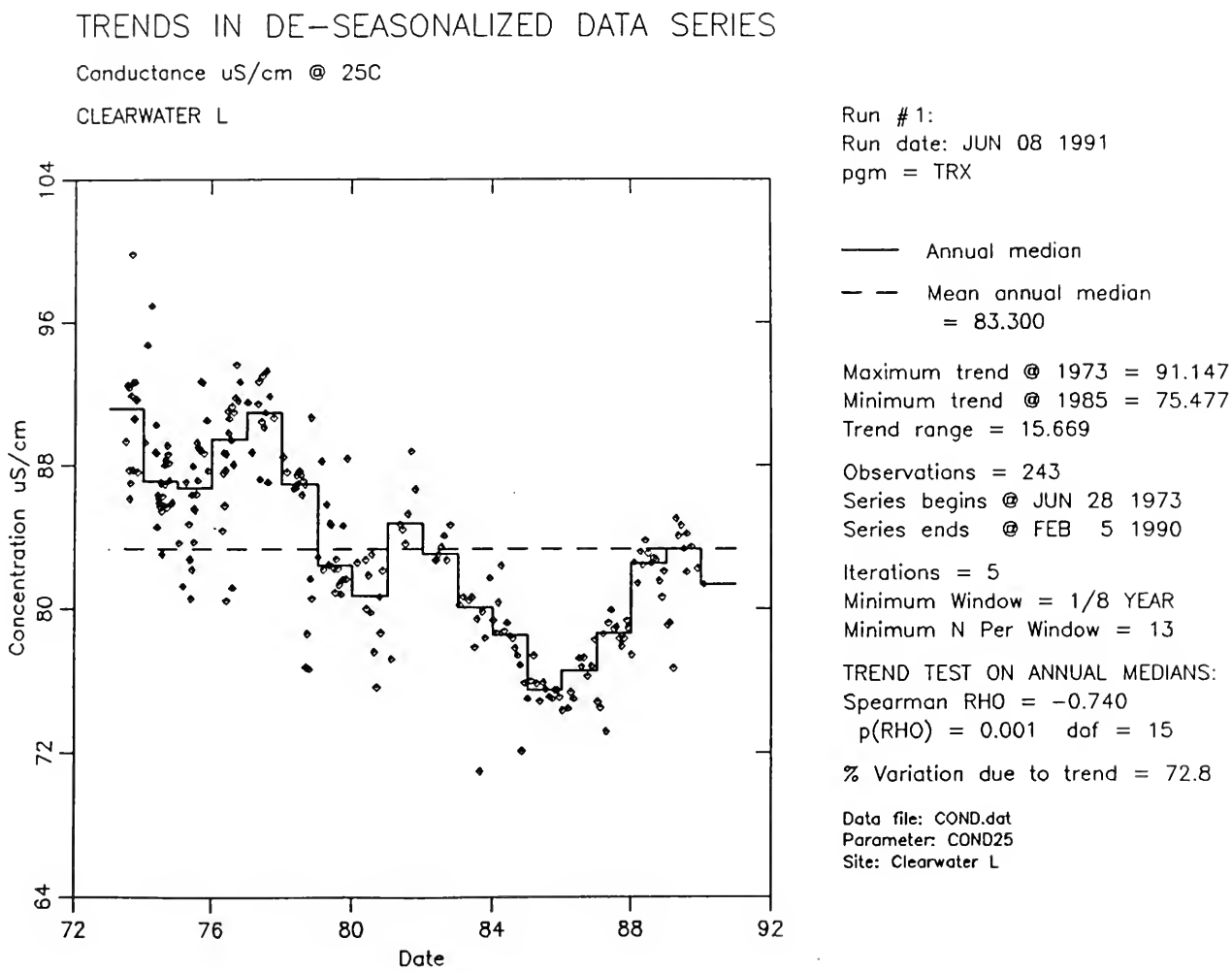
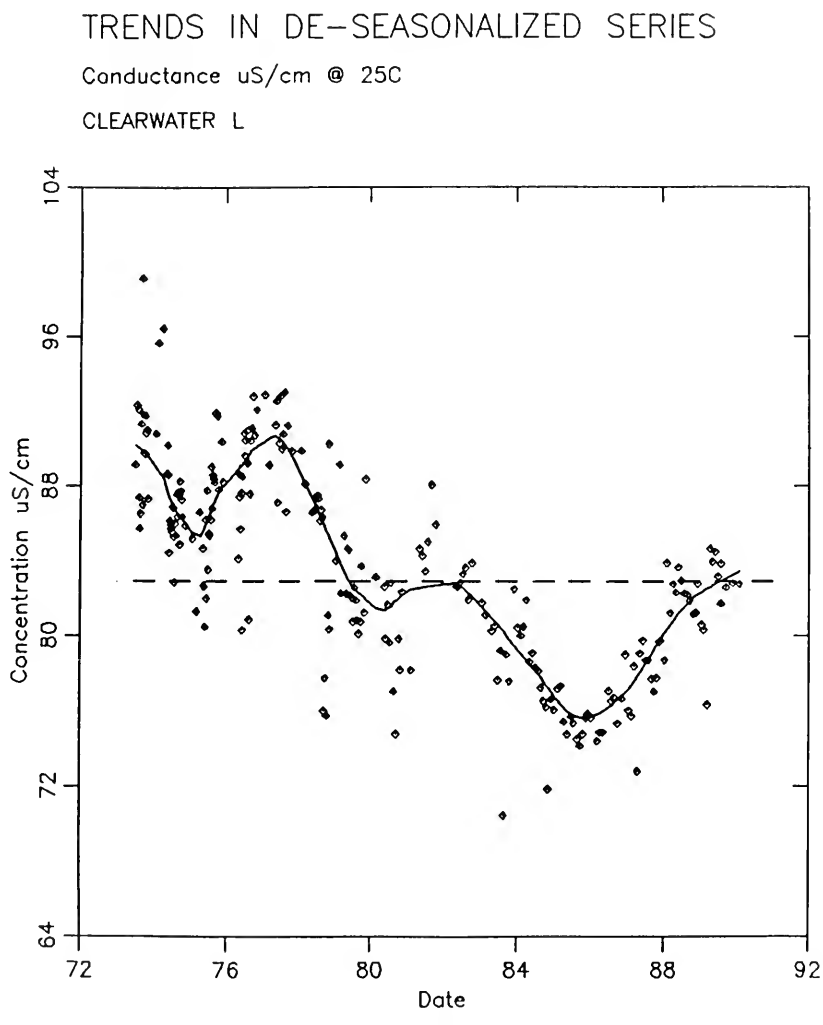


**Figure 5.1 Conductance trends: Filter 1, USI = 2 years.**




Run #1:
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 82.898

Maximum trend = 90.654
@ MAY 2 1977

Minimum trend = 75.620
@ OCT 22 1985

Trend range = 15.033

Observations = 243
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 1
Unit smoothing interval = 2 yr
Minimum smoothing on 35 observations

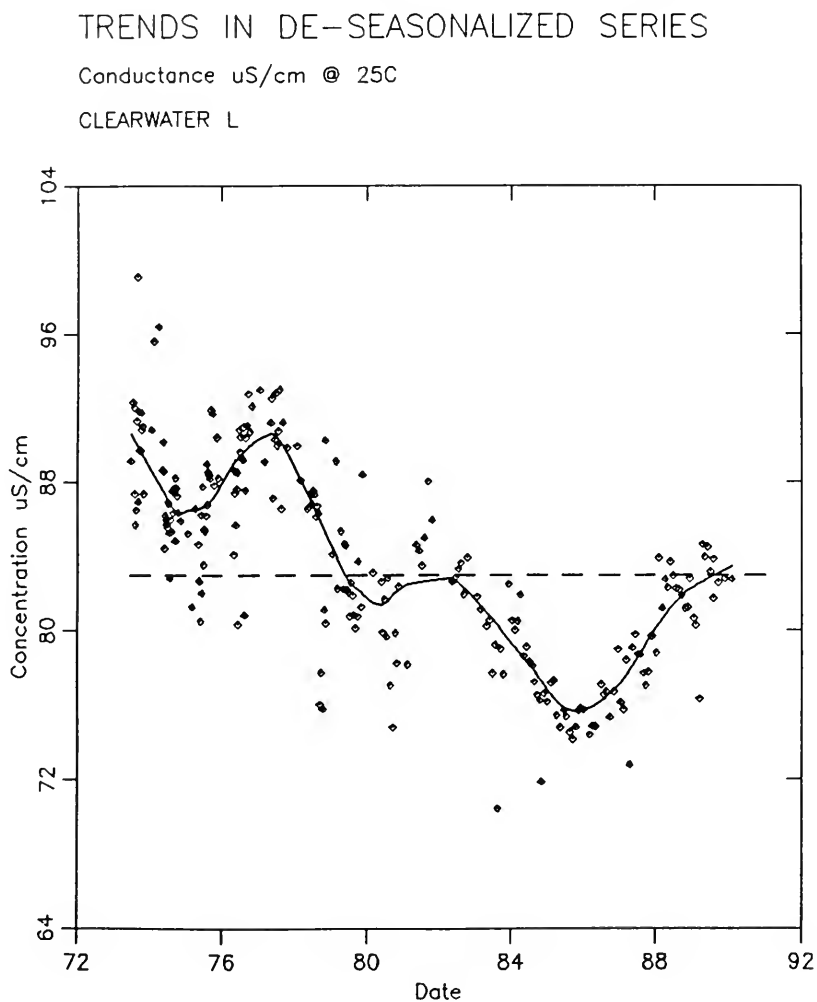
Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.789
p(RHO) = 0.000 dof = 15

% Variation due to trend = 75.7

Data file: COND.dat
Parameter: COND25
Site: Clearwater L

**Figure 5.2 Conductance trends: Filter 2, USI = 2 years.**




Run #1:
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 82.912

Maximum trend = 90.635
@ MAY 2 1977

Minimum trend = 75.621
@ OCT 22 1985

Trend range = 15.014

Observations = 243
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 2
Unit smoothing interval = 2 yr
Minimum smoothing on 35 observations

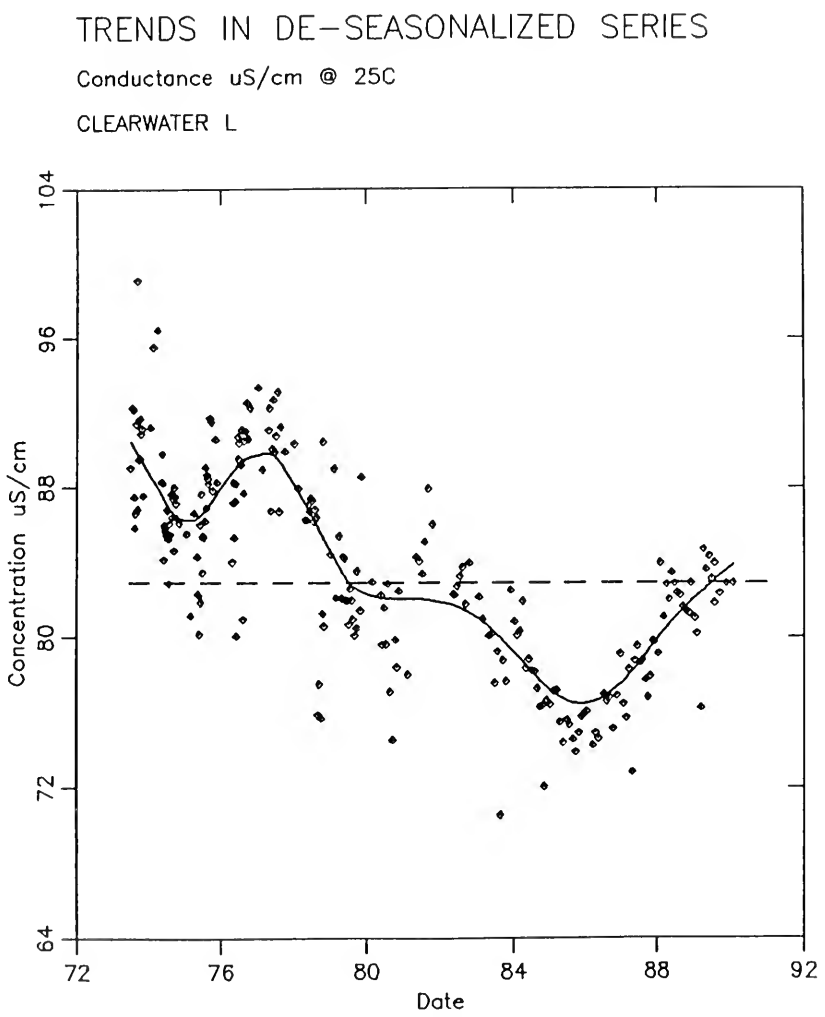
Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.806
p(RHO) = 0.000 dof = 15

% Variation due to trend = 74.6

Data file: COND.dat
Parameter: COND25
Site: Clearwater L

**Figure 5.3 Conductance trends: Filter 3, USI = 2 years.**




Run #1:
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 82.863

Maximum trend = 90.454
@ JUN 28 1973

Minimum trend = 76.450
@ DEC 12 1985

Trend range = 14.004

Observations = 243
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 3
Unit smoothing interval = 2 yr
Minimum smoothing on 35 observations

Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.814
p(RHO) = 0.000 dof = 15

% Variation due to trend = 73.4

Data file: COND.dat
Parameter: COND25
Site: Clearwater L

**Figure 5.4 Conductance trends: Filter 4, USI = 2 years.**

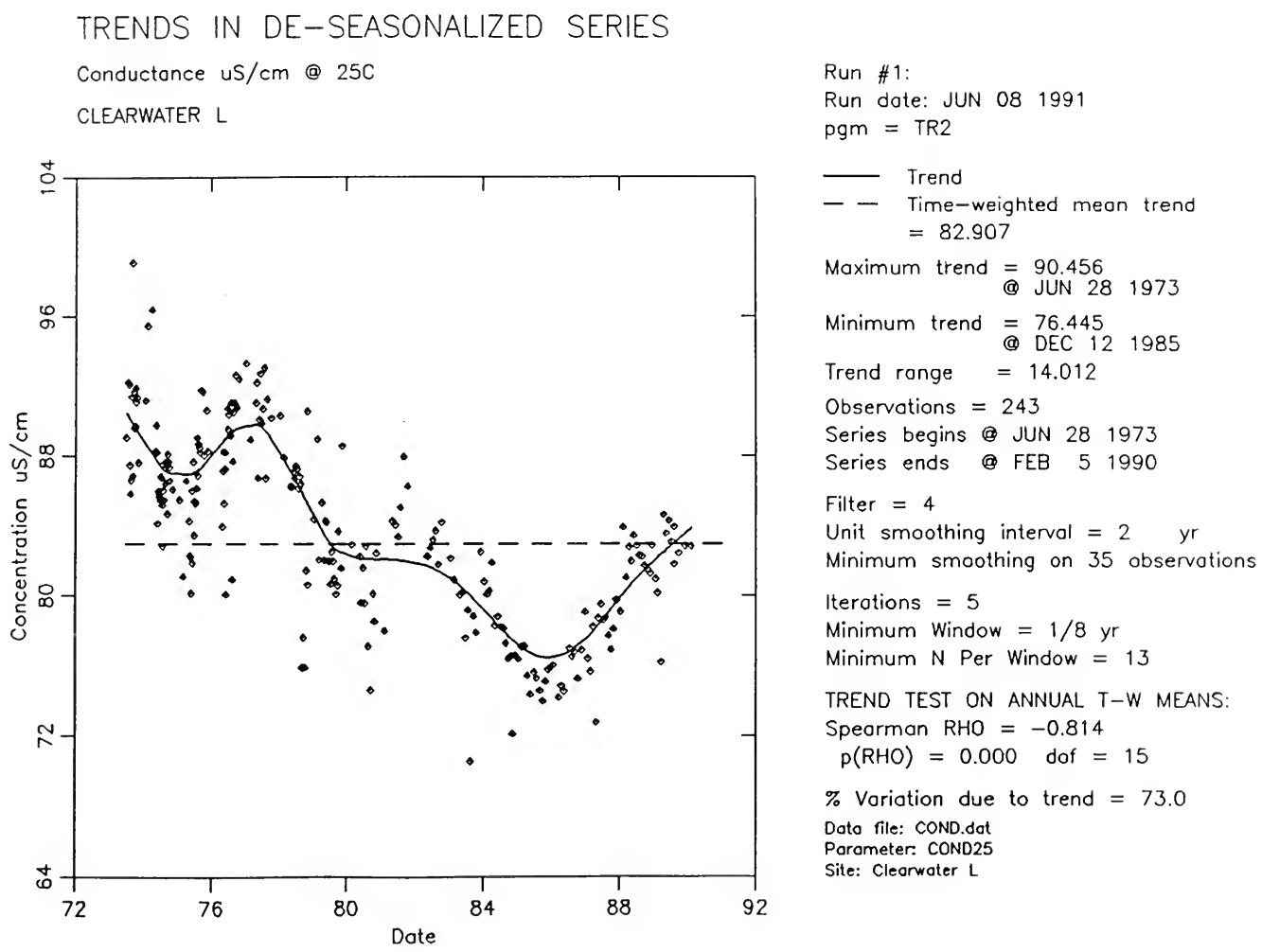
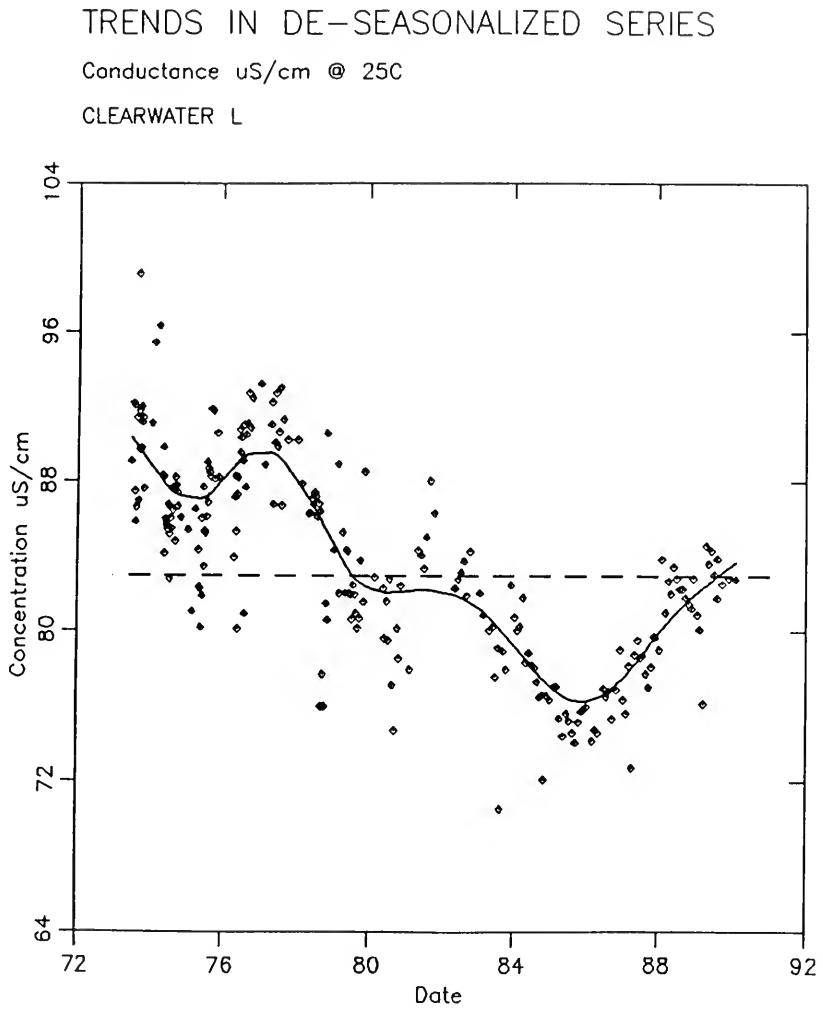


**Figure 5.5 Conductance trends: Filter 5, USI = 2 years.**




Run #1:
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 82.916

Maximum trend = 90.363
@ JUN 28 1973

Minimum trend = 76.266
@ DEC 12 1985

Trend range = 14.096

Observations = 243
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 5
Unit smoothing interval = 2 yr
Minimum smoothing on 35 observations

Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.811
p(RHO) = 0.000 dof = 15

% Variation due to trend = 73.1

Data file: COND.dat
Parameter: COND25
Site: Clearwater L

**Figure 5.6 Conductance trends: Filter 6, USI = 2 years.**

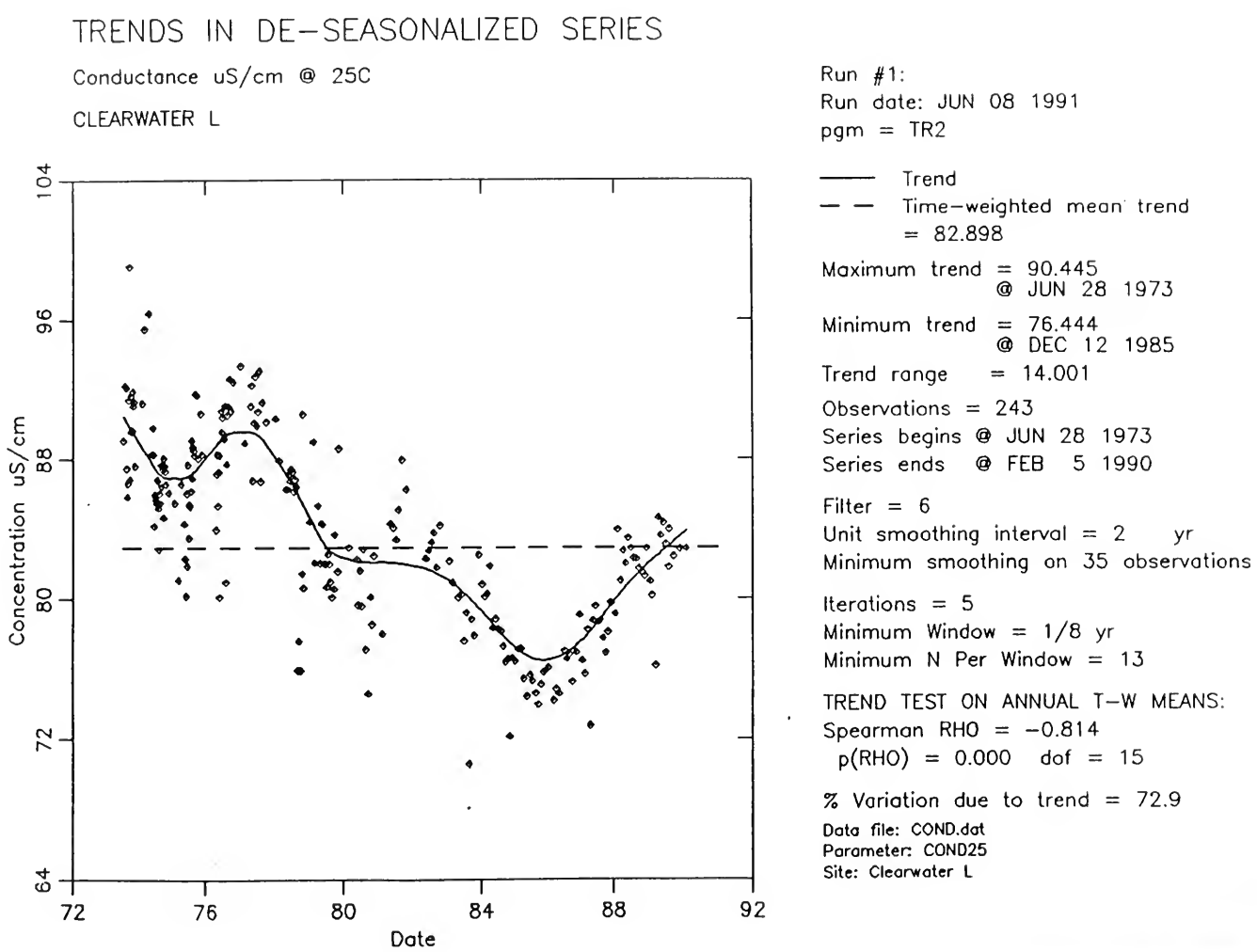
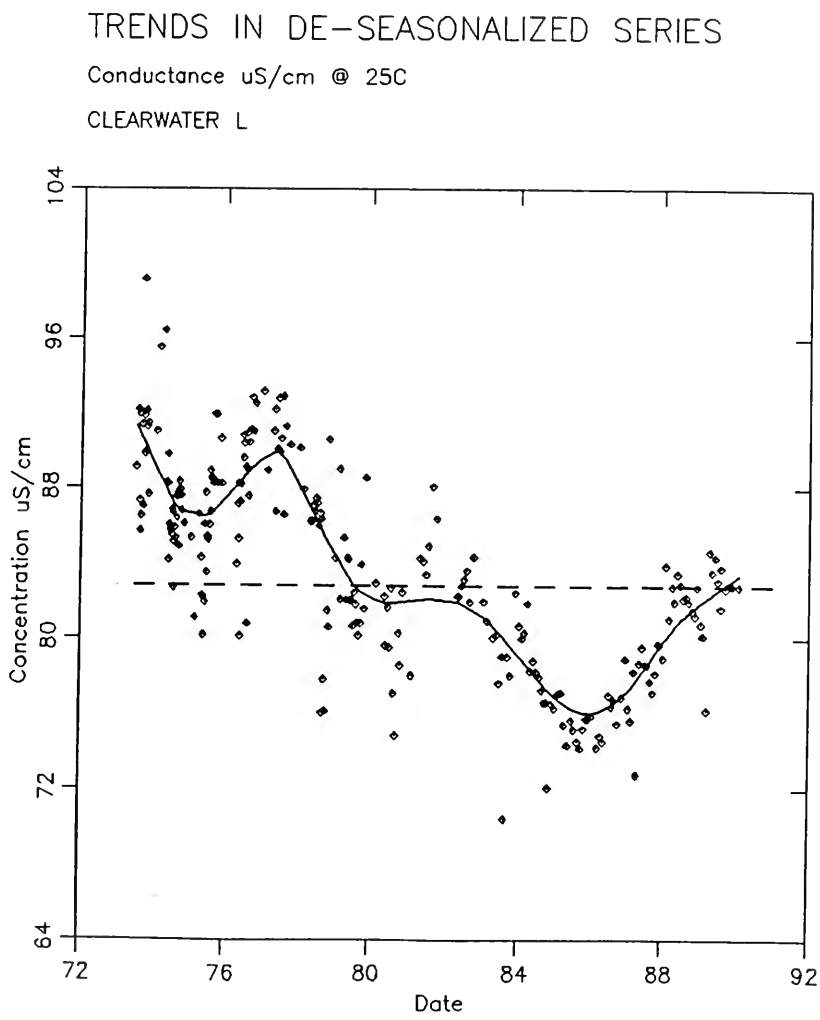


**Figure 5.7 Conductance trends: Filter 7, USI = 2 years.**




Run #1:
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 82.758

Maximum trend = 91.309
@ JUN 28 1973

Minimum trend = 76.056
@ JAN 15 1986

Trend range = 15.254

Observations = 243
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 7
Unit smoothing interval = 2 yr
Minimum smoothing on 35 observations

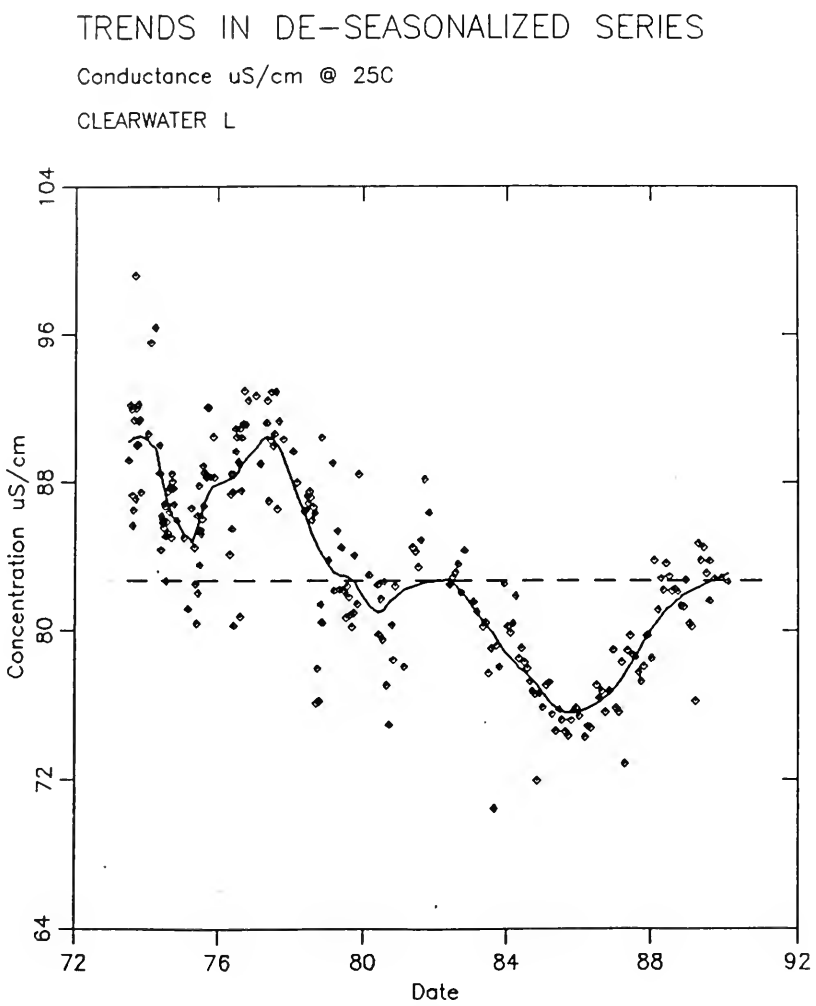
Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.811
p(RHO) = 0.000 dof = 15

% Variation due to trend = 74.6

Data file: COND.dat
Parameter: COND25
Site: Clearwater L

**Figure 5.8 Conductance trends: Filter 9, USI = 2 years.**




Run #1:
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 82.686

Maximum trend = 90.460
@ MAY 2 1977

Minimum trend = 75.569
@ OCT 22 1985

Trend range = 14.891

Observations = 243
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 9
Unit smoothing interval = 2 yr
Minimum smoothing on 35 observations

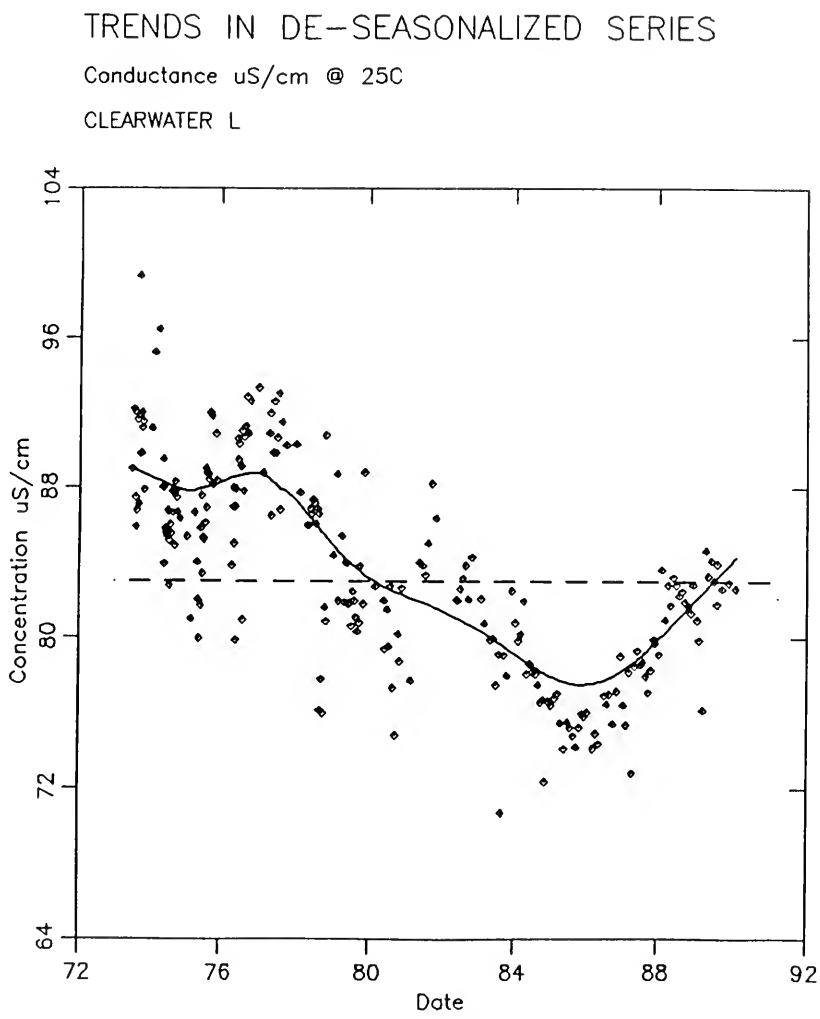
Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.806
p(RHO) = 0.000 dof = 15

% Variation due to trend = 77.7

Data file: COND.dat
Parameter: COND25
Site: Clearwater L

**Figure 6.1 Conductance trends: Filter 4, USI = 3 years.**




Run #1:
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend = 83.010

Maximum trend = 89.004 @ JUN 28 1973

Minimum trend = 77.508 @ OCT 22 1985

Trend range = 11.496

Observations = 243
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 4
Unit smoothing interval = 3 yr
Minimum smoothing on 53 observations

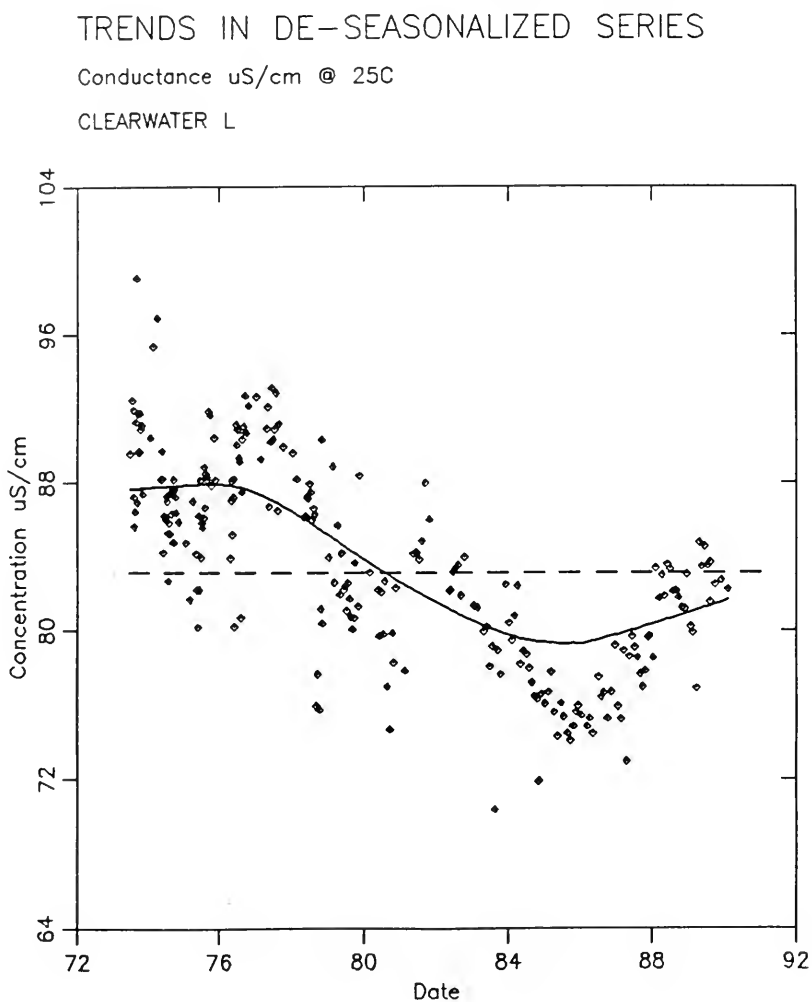
Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.801
p(RHO) = 0.000 dof = 15

% Variation due to trend = 67.8

Data file: COND.dat
Parameter: COND25
Site: Clearwater L

**Figure 6.2 Conductance trends: Filter 4, USI = 5 years.**




Run #1:
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 83.080

Maximum trend = 87.901
@ NOV 27 1975

Minimum trend = 79.269
@ JUN 26 1985

Trend range = 8.632

Observations = 243
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 4
Unit smoothing interval = 5 yr
Minimum smoothing on 88 observations

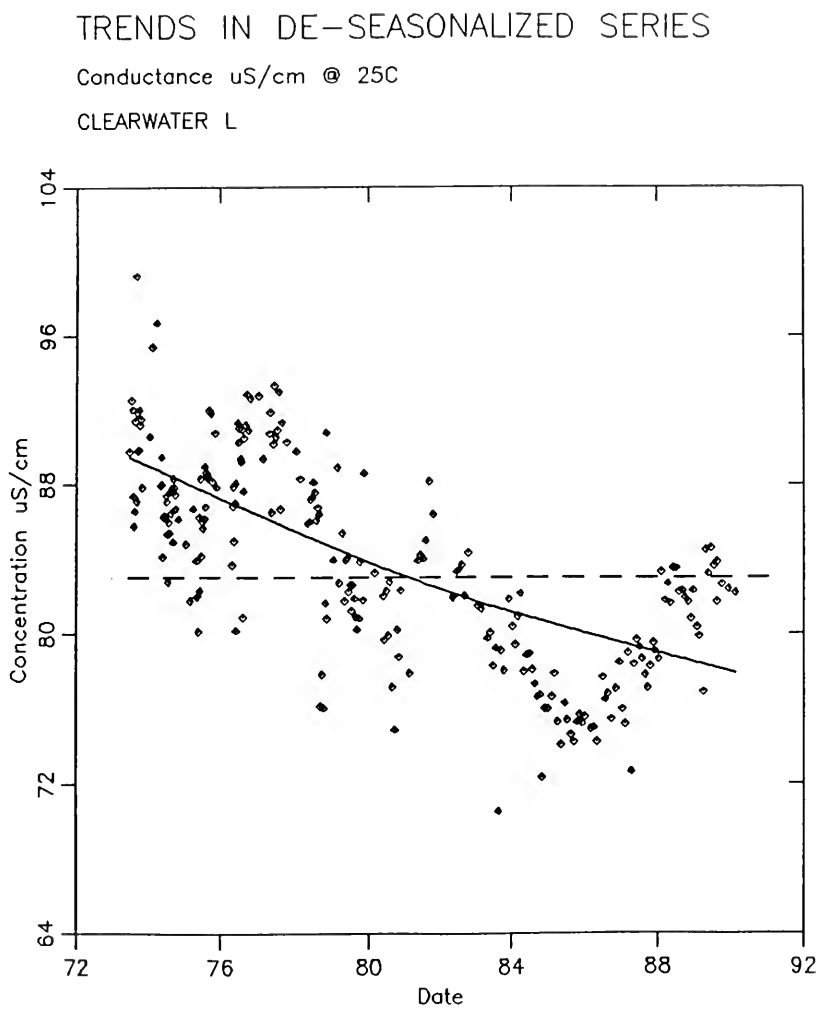
Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.860
p(RHO) = 0.000 dof = 15

% Variation due to trend = 59.8

Data file: COND.dat
Parameter: COND25
Site: Clearwater L

**Figure 6.3 Conductance trends: Filter 4, USI = 10 years.**




Run #1:
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 82.976

Maximum trend = 89.450
@ JUN 28 1973

Minimum trend = 77.888
@ FEB 5 1990

Trend range = 11.562

Observations = 243
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 4
Unit smoothing interval = 10 yr
Minimum smoothing on 175observations

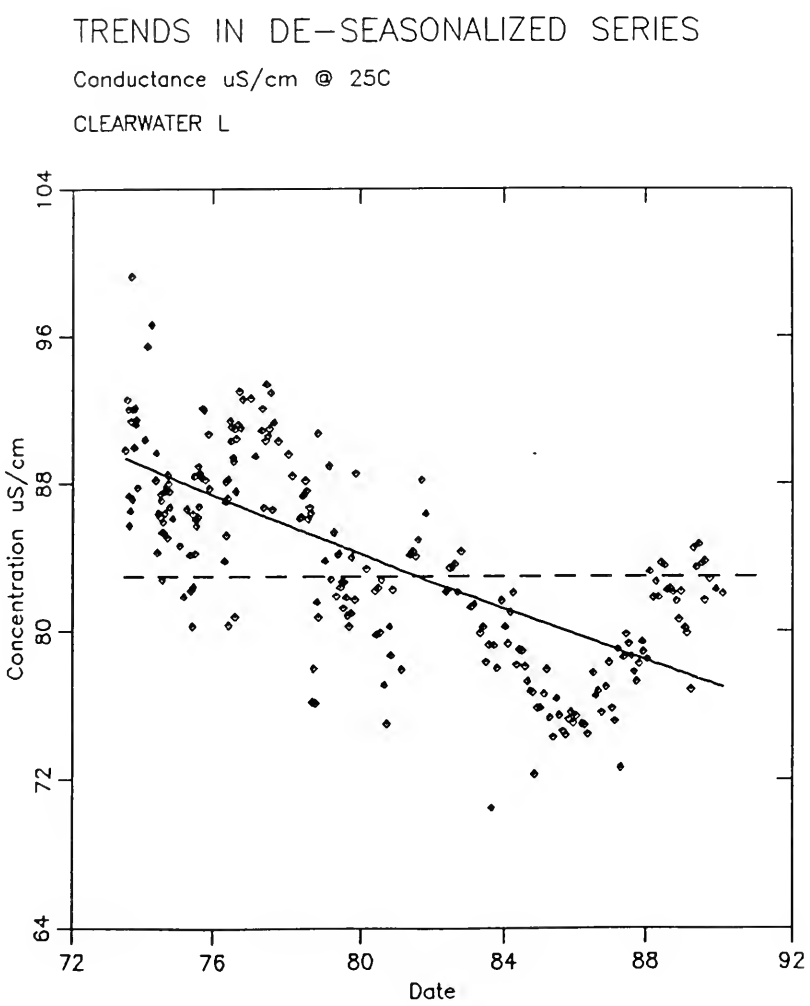
Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −1.000
p(RHO) = 0.000 dof = 15

% Variation due to trend = 52.6

Data file: COND.dat
Parameter: COND25
Site: Clearwater L

**Figure 6.4 Conductance trends: Filter 4, USI = 16.6 years.**




Run #1:
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 82.939

Maximum trend = 89.359
@ JUN 28 1973

Minimum trend = 76.961
@ FEB 5 1990

Trend range = 12.398

Observations = 243
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 4
Unit smoothing interval = 16.60yr
Minimum smoothing on 243observations

Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −1.000
p(RHO) = 0.000 dof = 15

% Variation due to trend = 51.6

Data file: COND.dat
Parameter: COND25
Site: Clearwater L

**Figure 7.1 Time series exploration with TR1.**

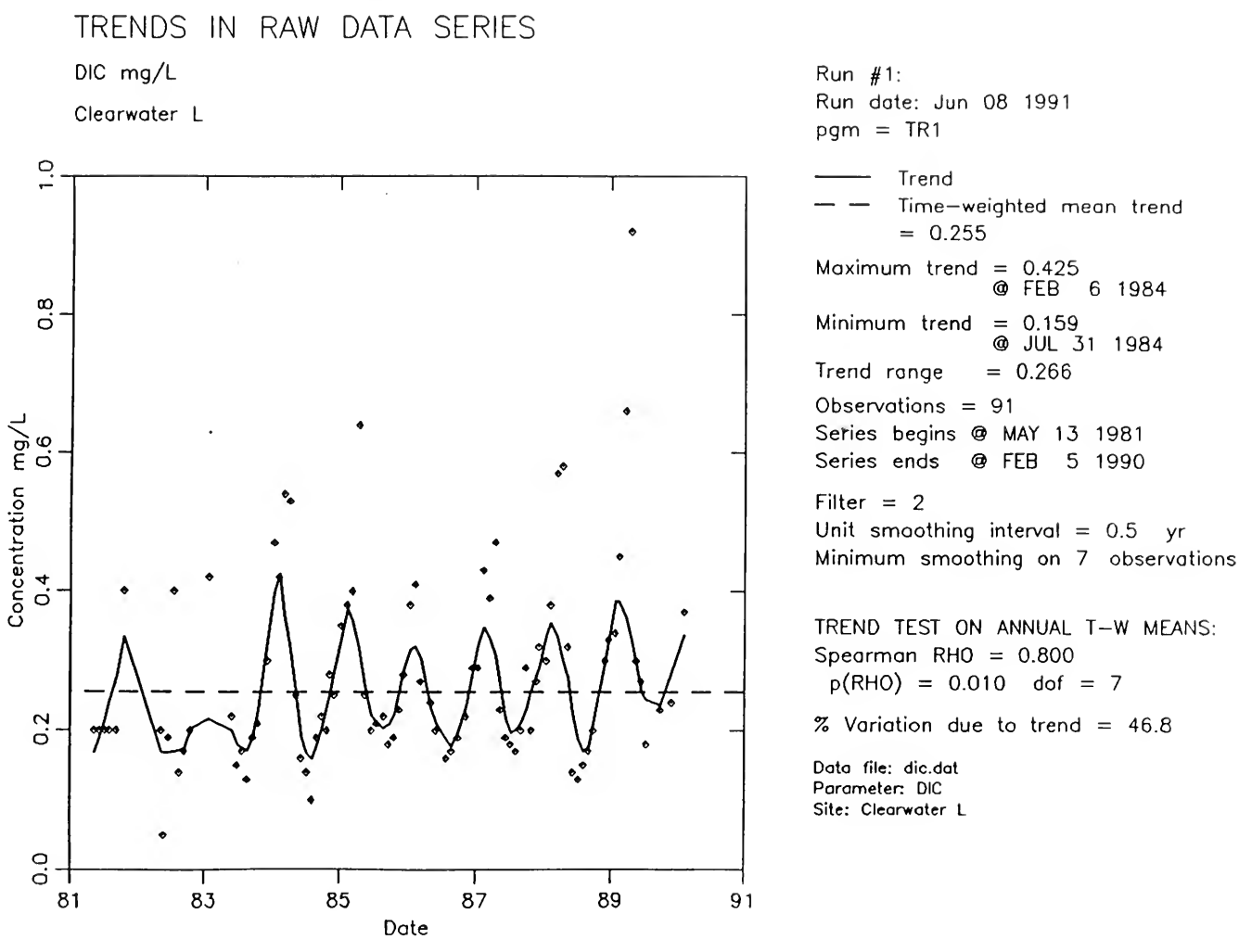


**Figure 7.2 TR1 applied to very small data sets.**

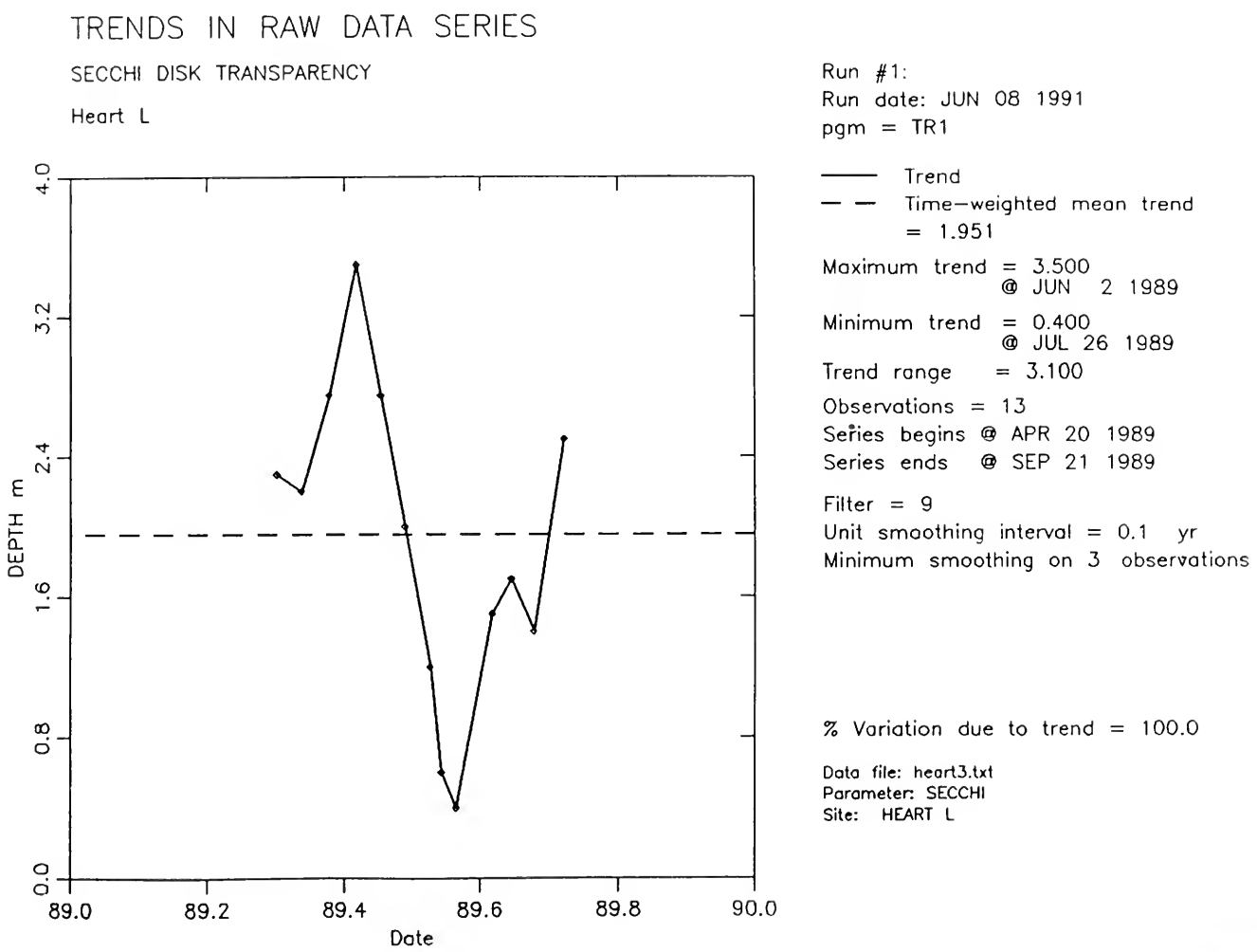


**Figure 8.1 Seasonal trends in DIC, Clearwater L.**

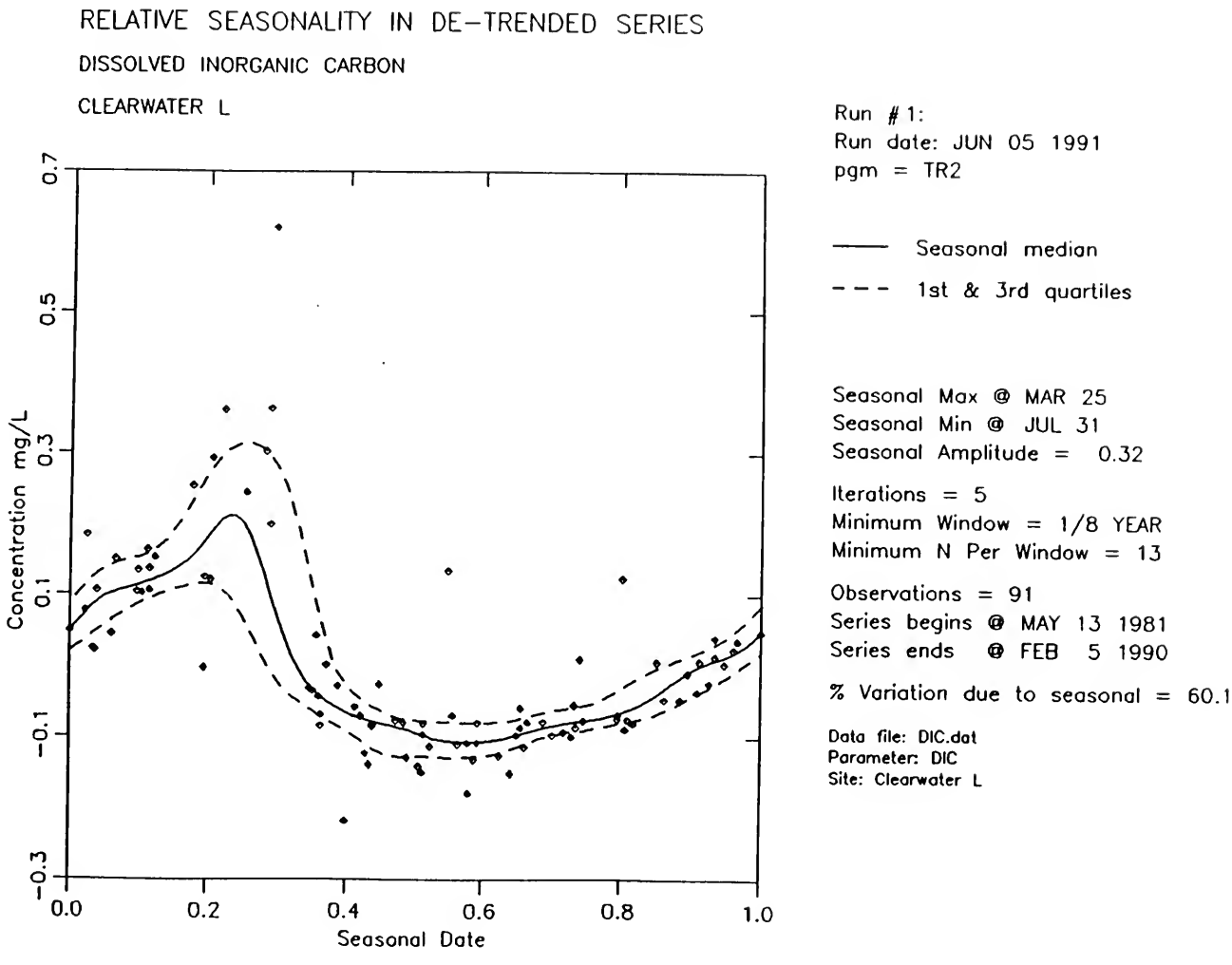


**Figure 8.2 Seasonal trends in conductance, Clearwater L.**

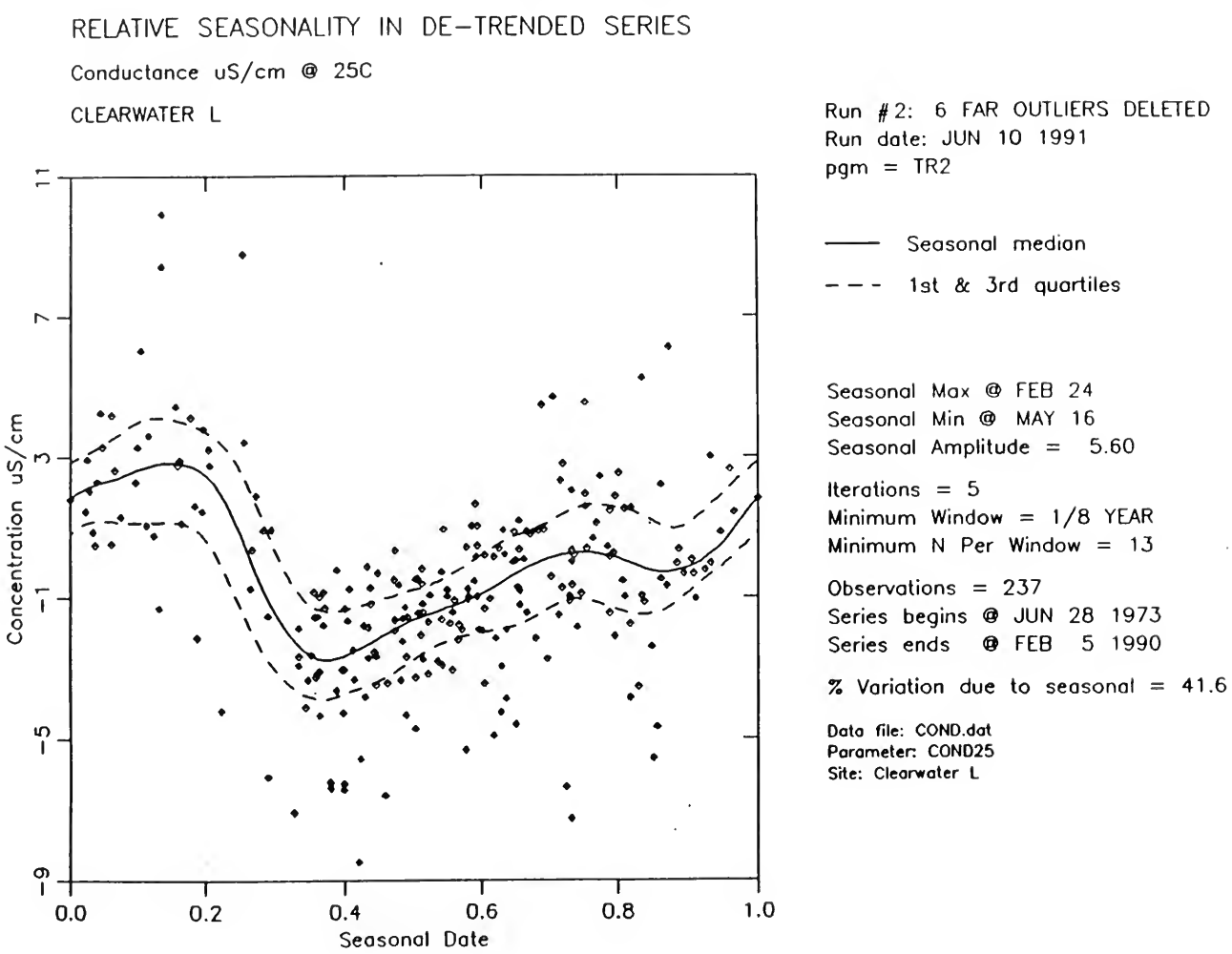


**Figure 8.3 Seasonal trends in manganese, Clearwater L.**

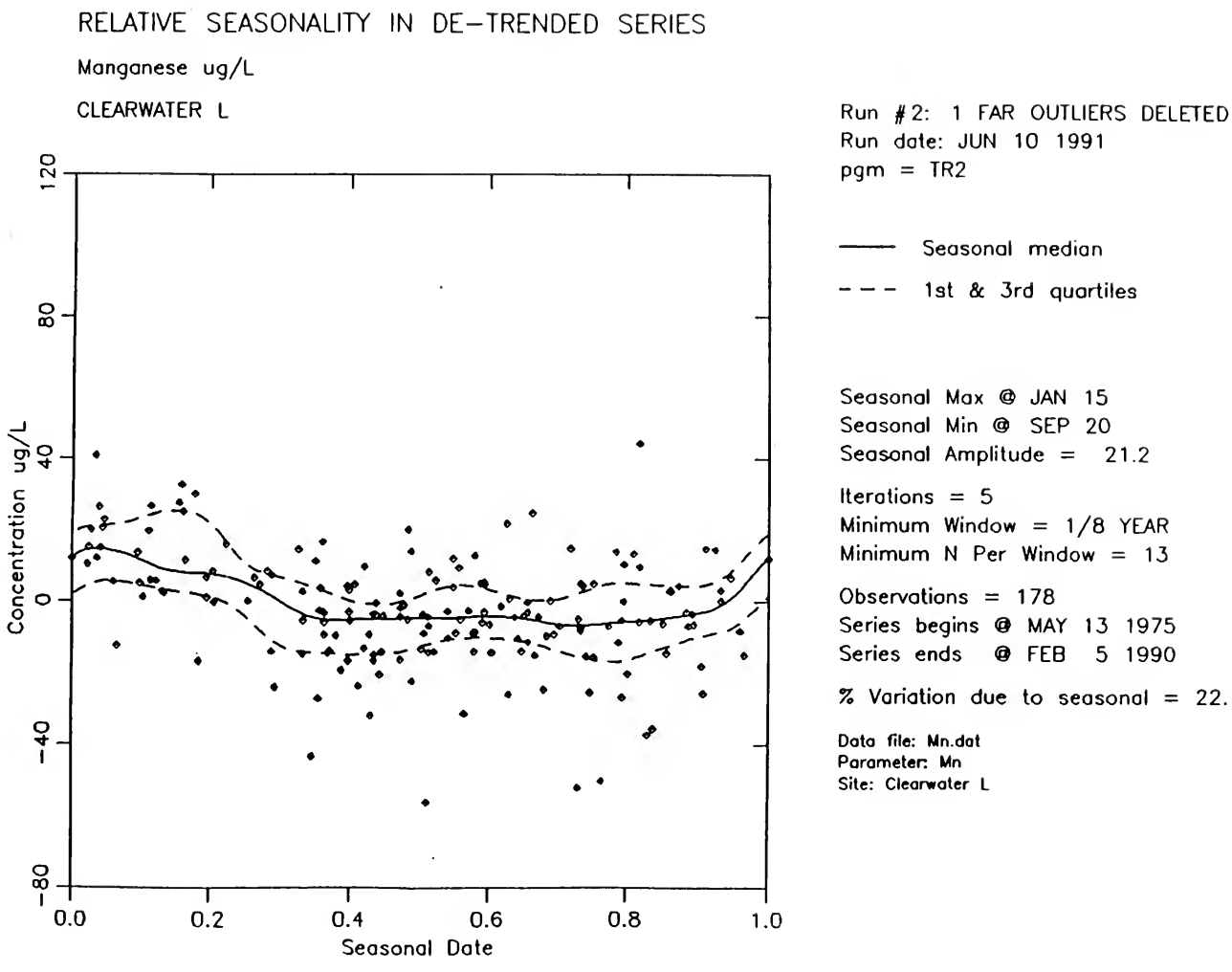


**Figure 8.4 Seasonal trends in potassium, Clearwater L.**

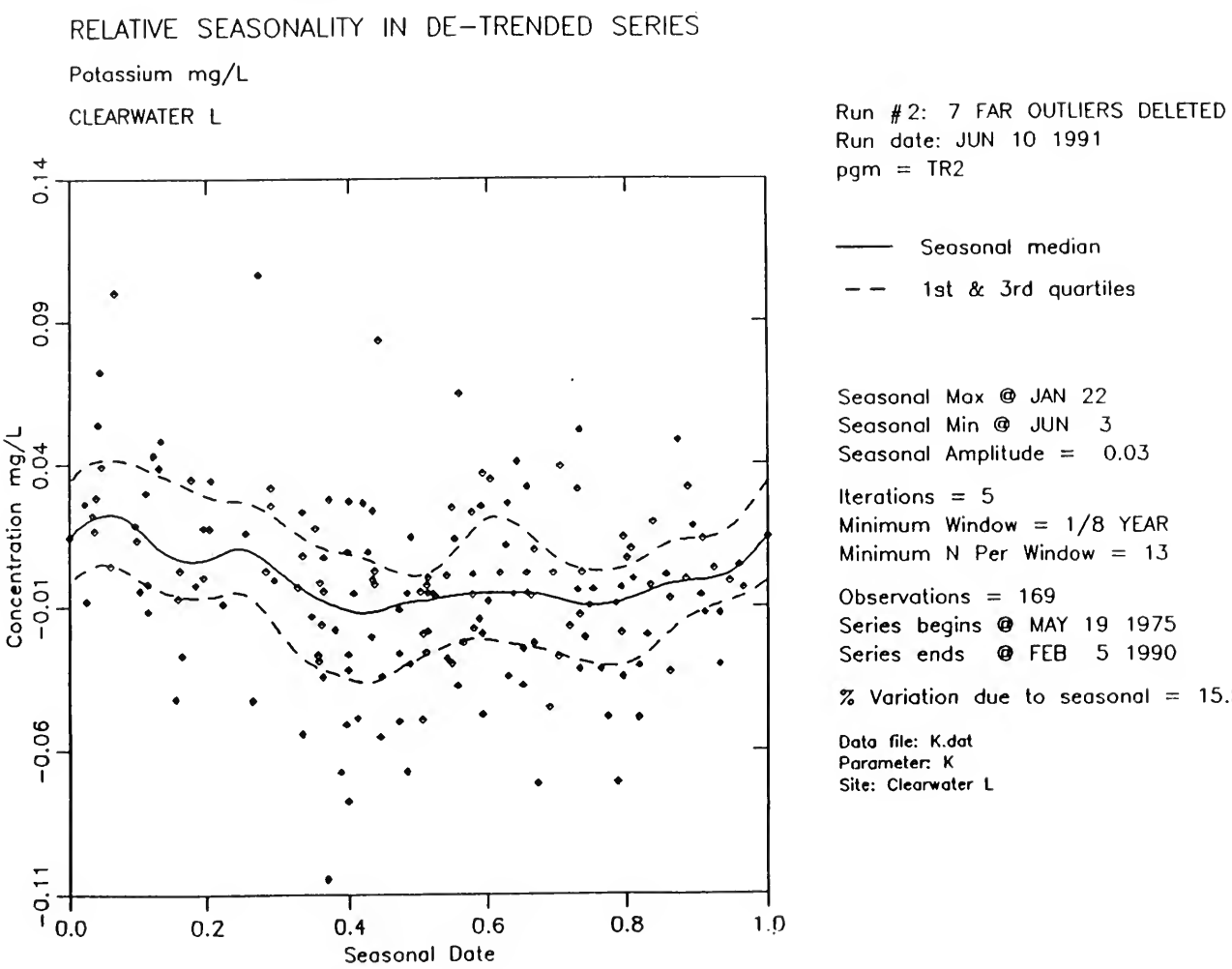


**Figure 8.5a** **Seasonal trends in aluminum, Clearwater L.**
**Control parameters:** $\hat{m} = 11$ **and** $h = 10$.

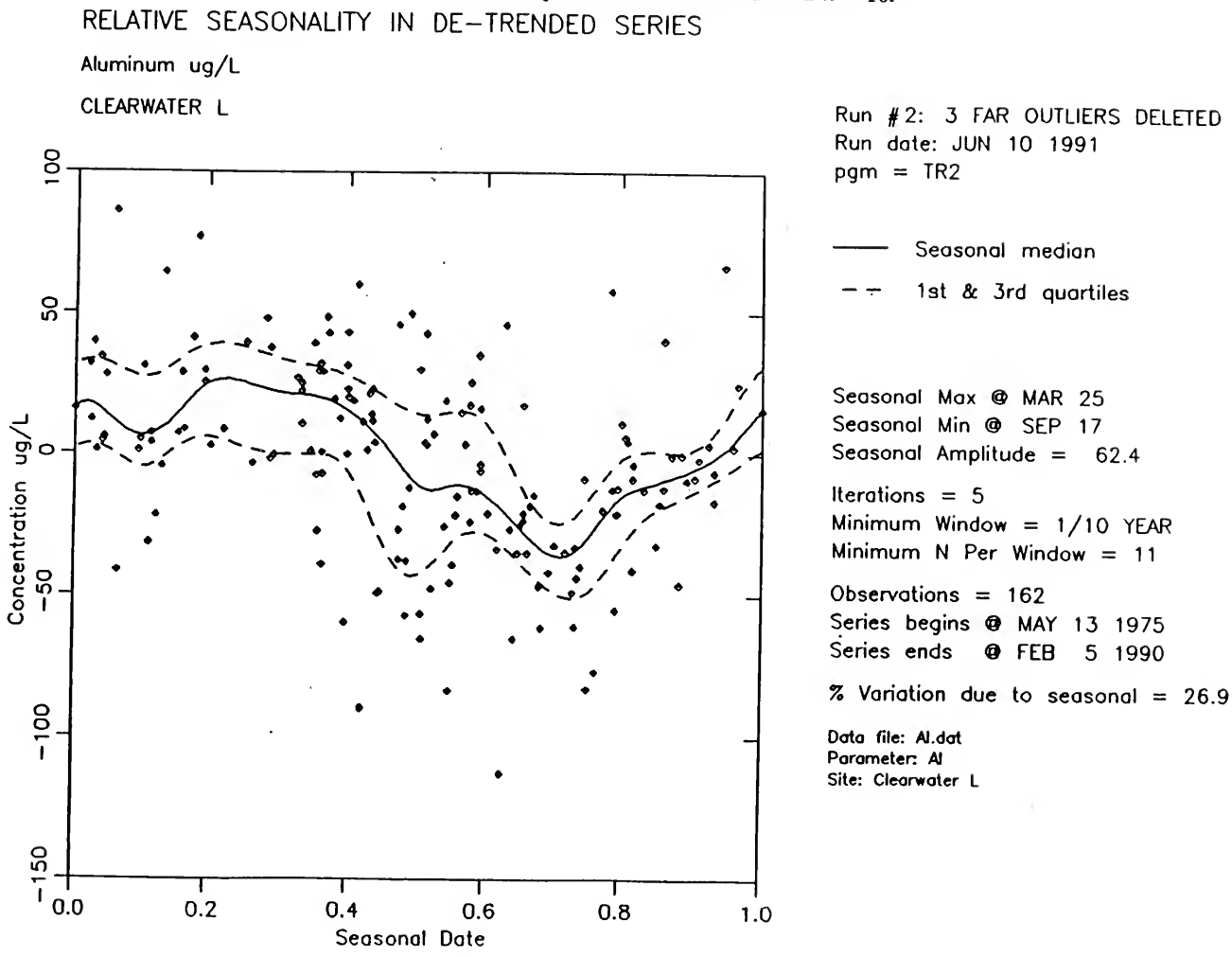


**Figure 8.5b** **Seasonal trends in aluminum, Clearwater L.**
**Control parameters:** $\hat{m} = 13$ **and** $h = 8$

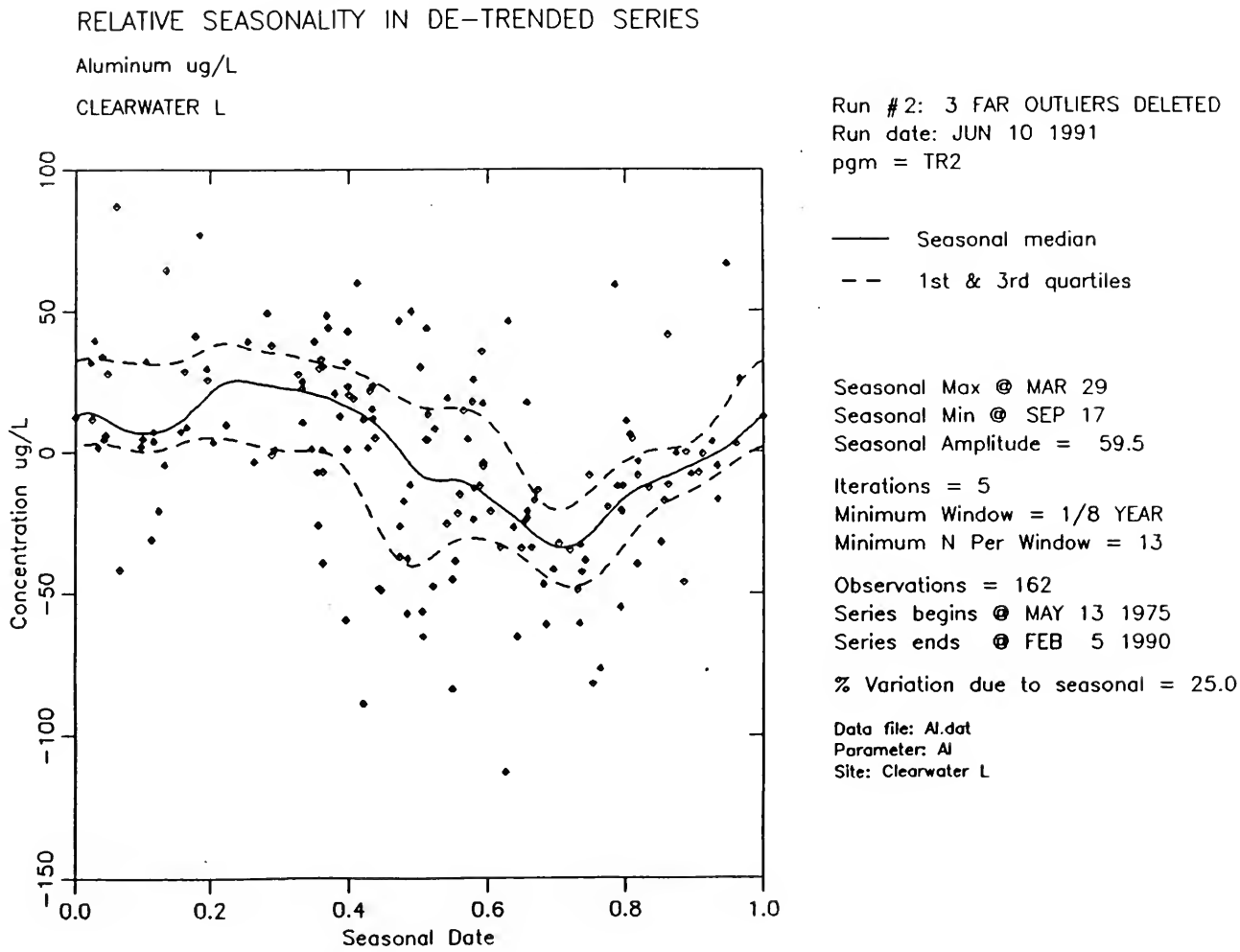


**Figure 8.5c Seasonal trends in aluminum, Clearwater L: Control parameters: $\hat{m} = 15$ and $h = 6$**

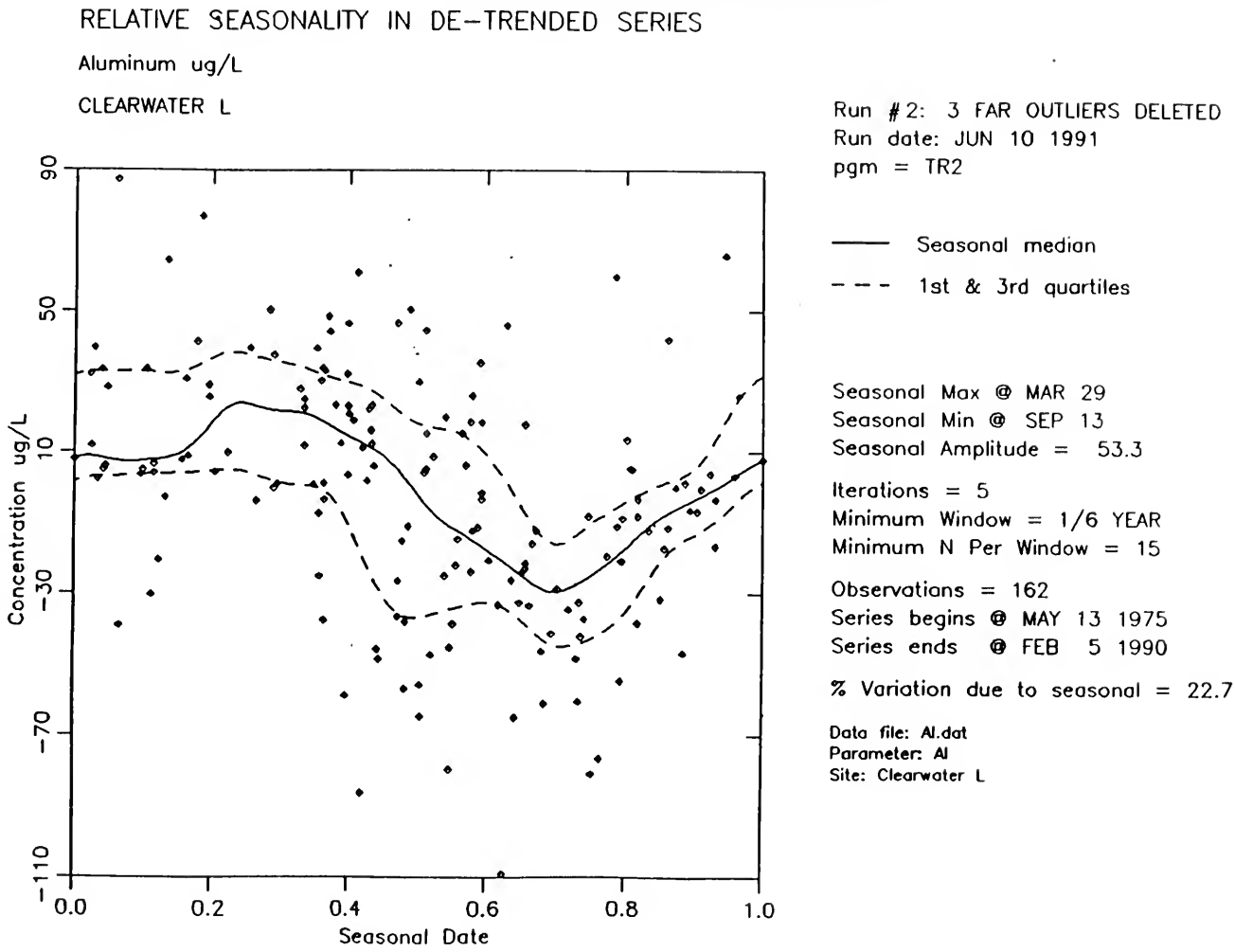


**Figure 8.5d Aluminum trends, Clearwater L:**

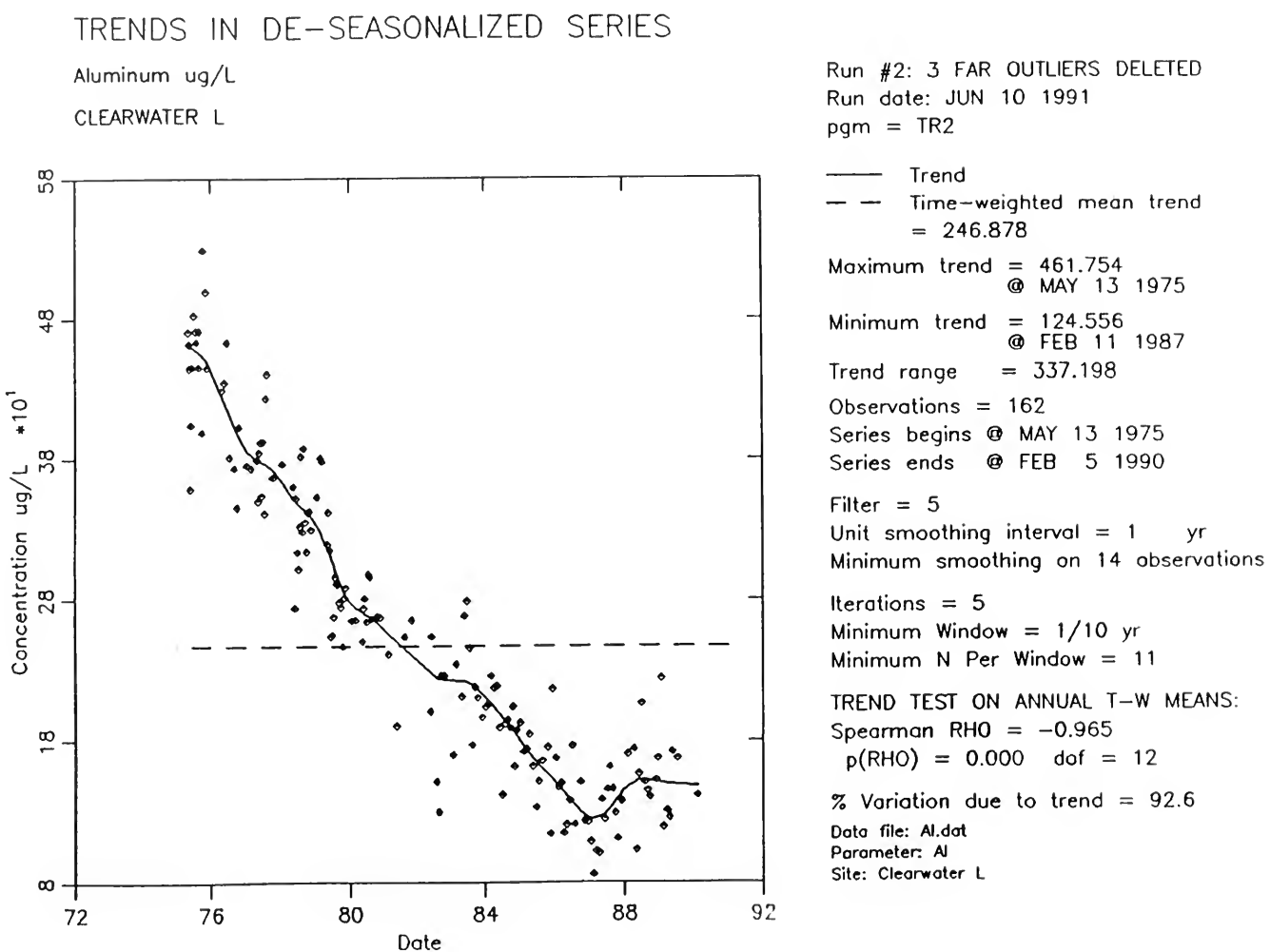


**Figure 8.6a Chloride Historical Trends Clearwater L: Program TRX**

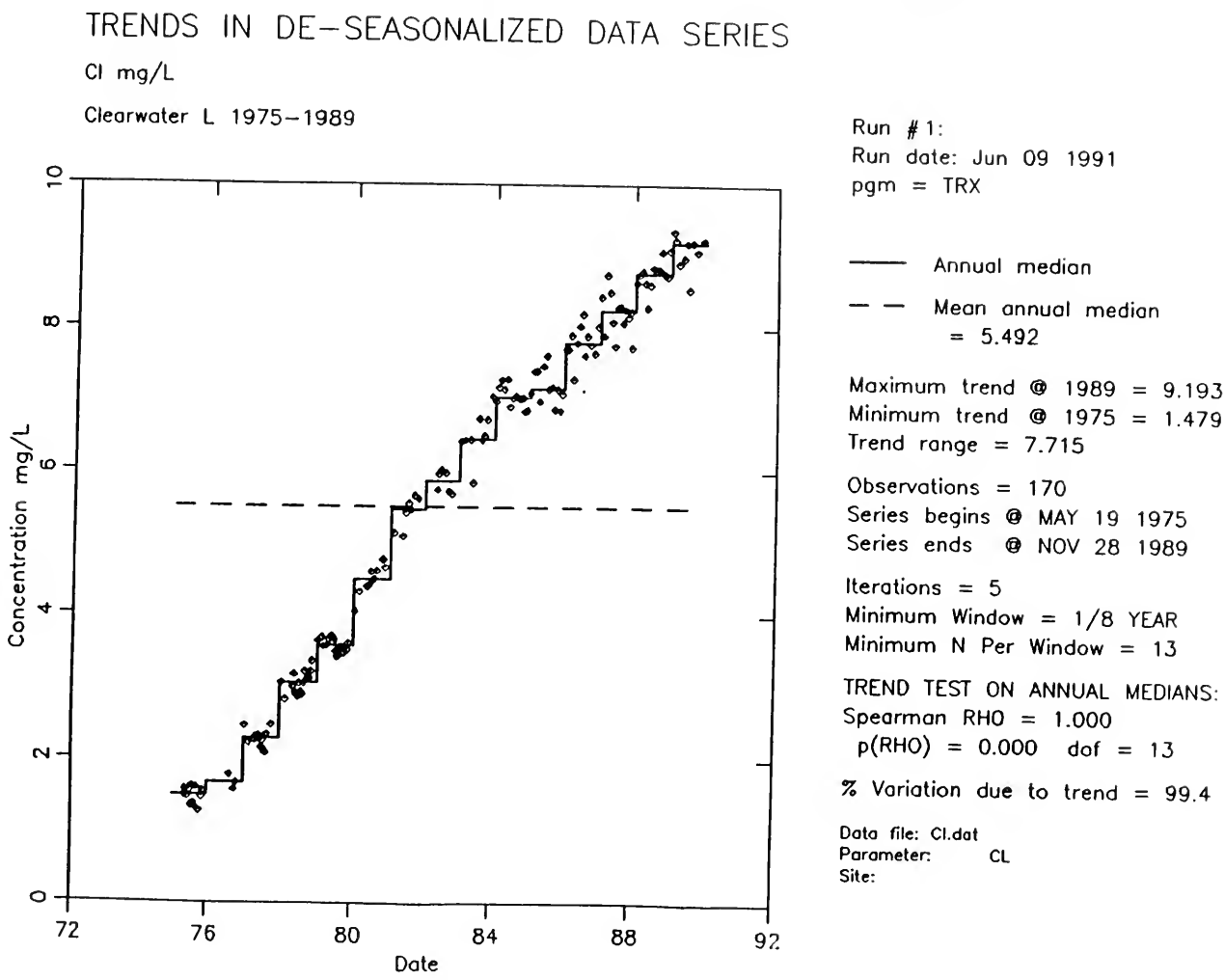


**Figure 8.6b Chloride Seasonal Trends Clearwater L: Program TRX**

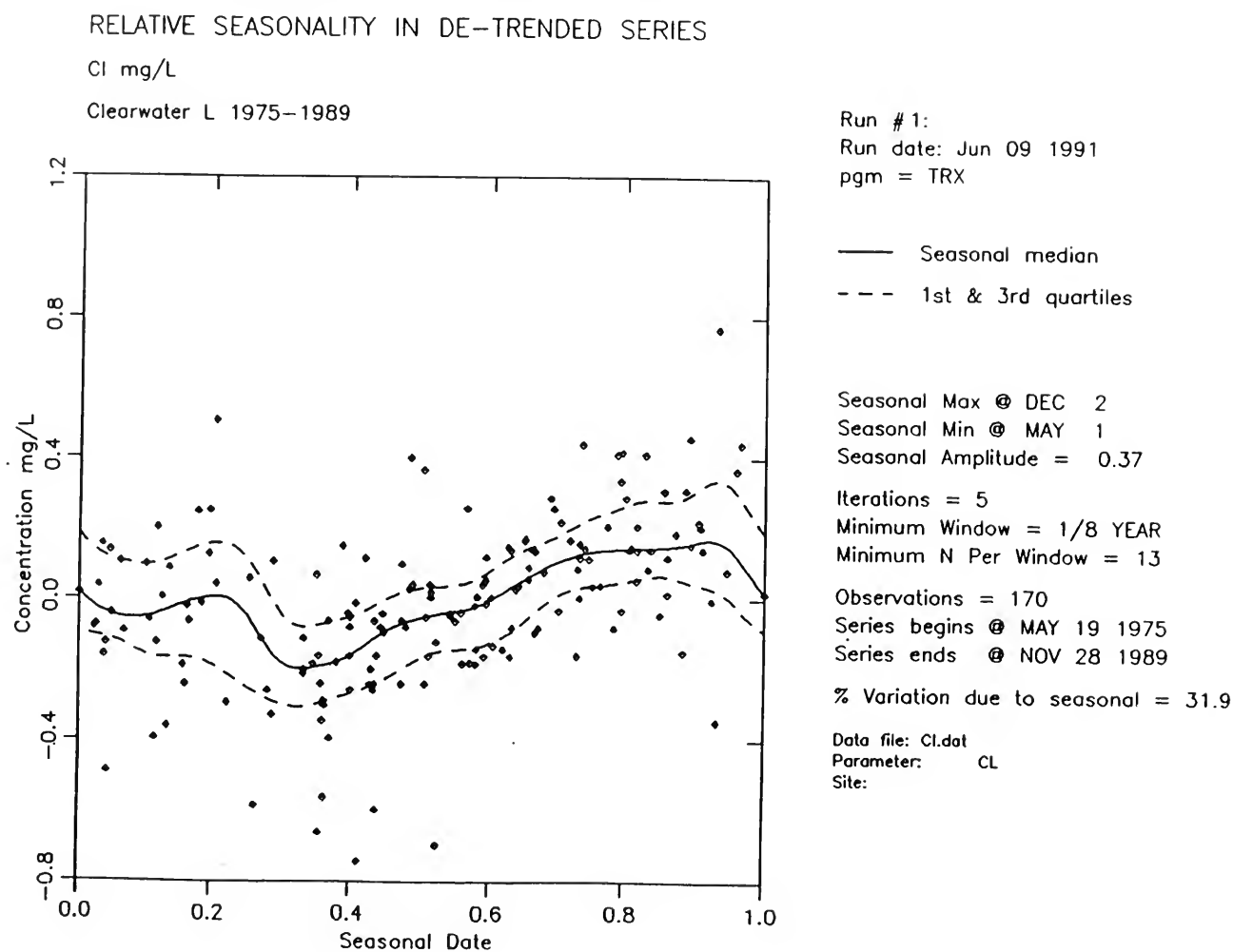
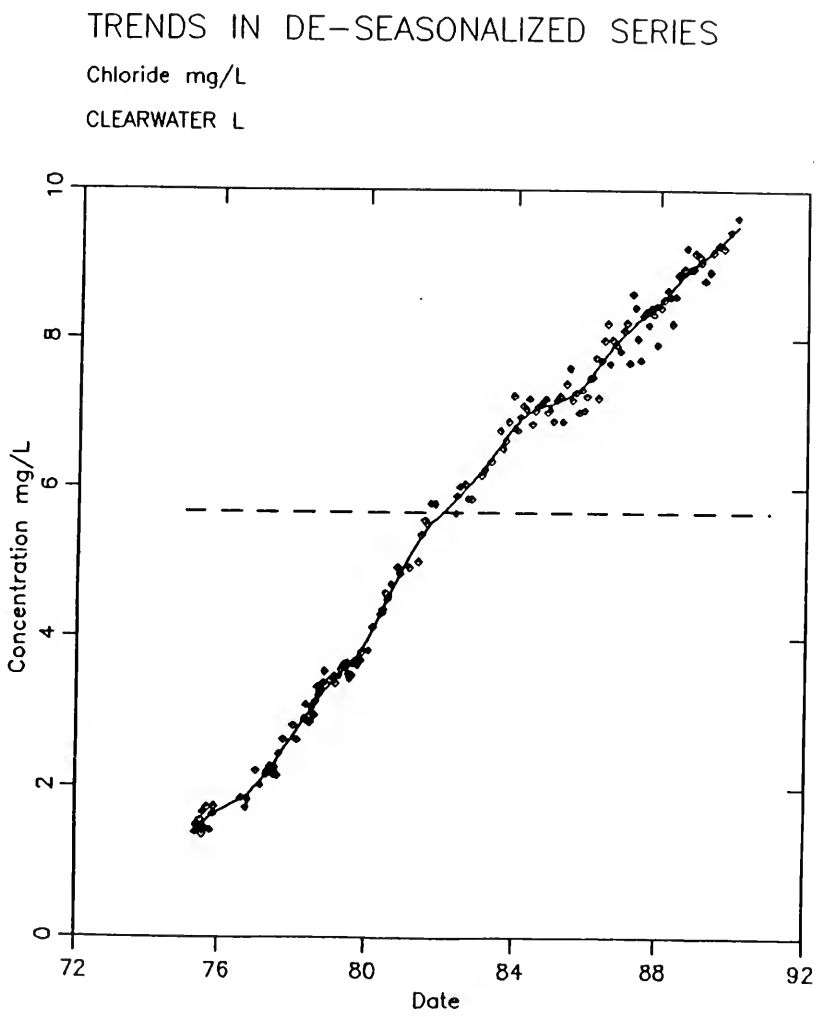


**Figure 8.6c Chloride Historical Trends Clearwater L: Program TR2**




Run #2: 2 FAR OUTLIERS DELETED
Run date: JUN 08 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 5.661

Maximum trend = 9.525
@ FEB 5 1990

Minimum trend = 1.446
@ MAY 19 1975

Trend range = 8.080

Observations = 169
Series begins @ MAY 19 1975
Series ends @ FEB 5 1990

Filter = 5
Unit smoothing interval = 1 yr
Minimum smoothing on 14 observations

Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = 1.000
p(RHO) = 0.000 dof = 12

% Variation due to trend = 99.6

Data file: Cl.dat
Parameter: CL
Site:

**Figure 8.6d Chloride Seasonal Trends Clearwater L: Program TR2**

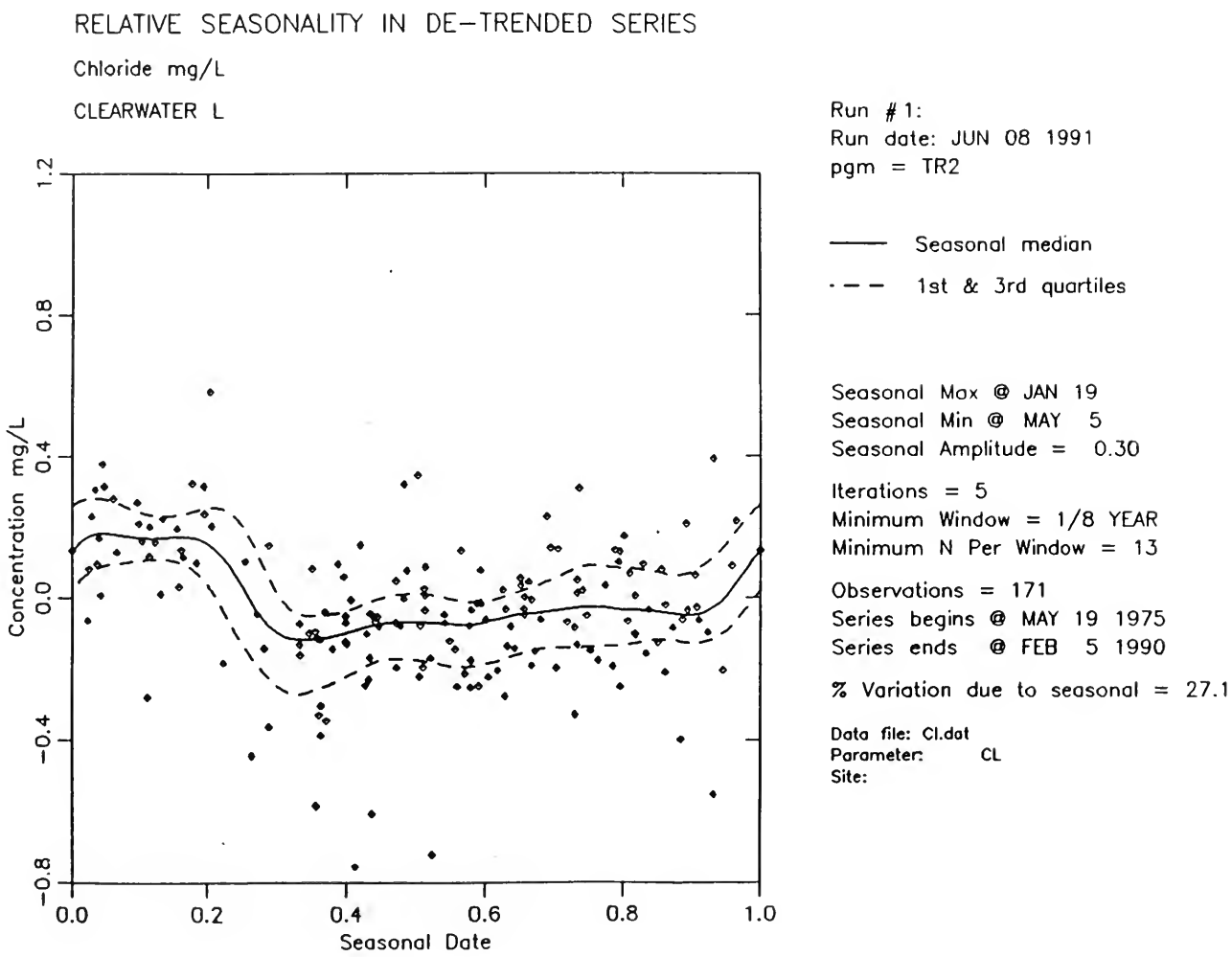


**Figure 9.a Seasonal trends in arsenic, Moira R.**
**Control parameters: $\hat{m} = 15$ and $h = 26$.**

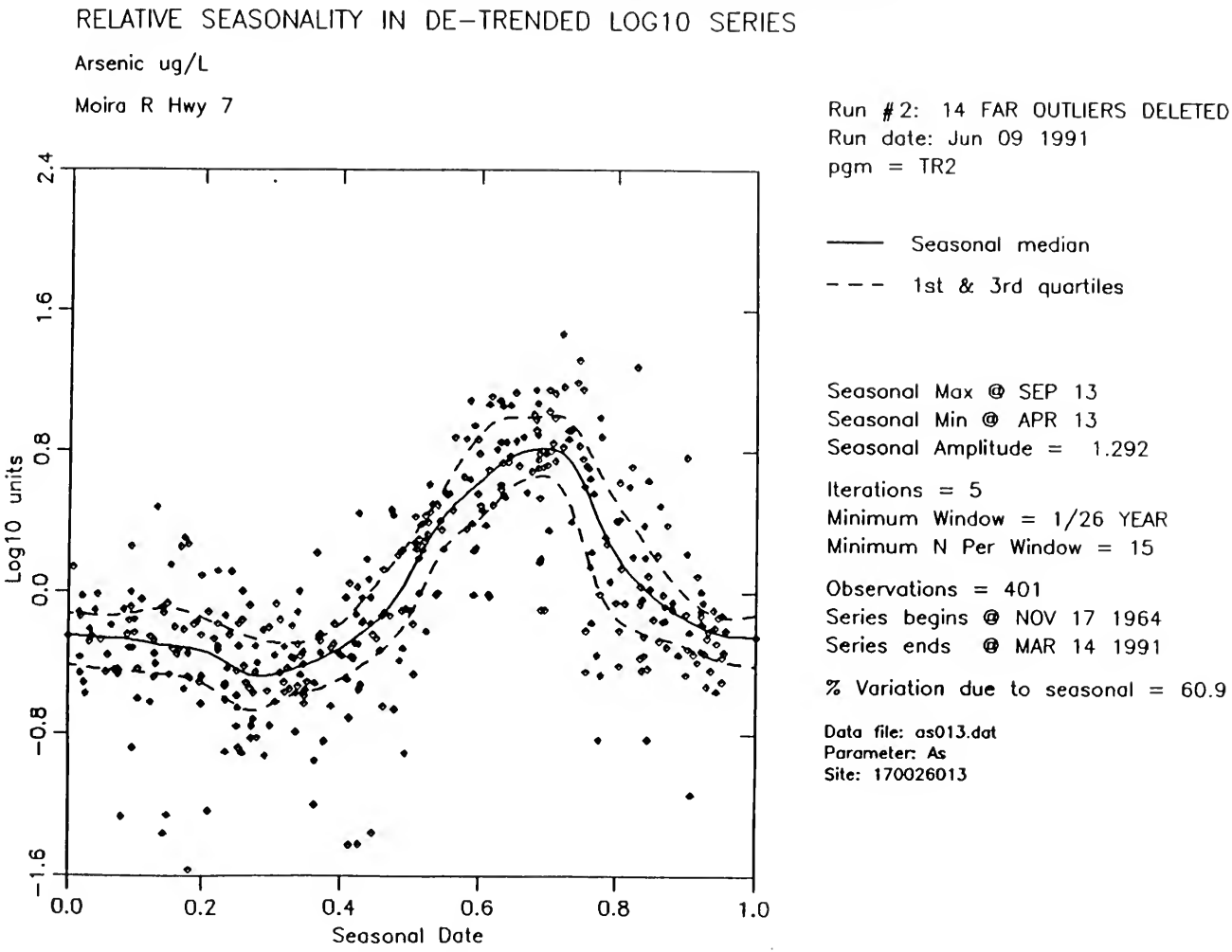


**Figure 9.b Seasonal trends in arsenic, Moira R.**
**Control parameters: $\widehat{m} = 35$ and $h = 12$**

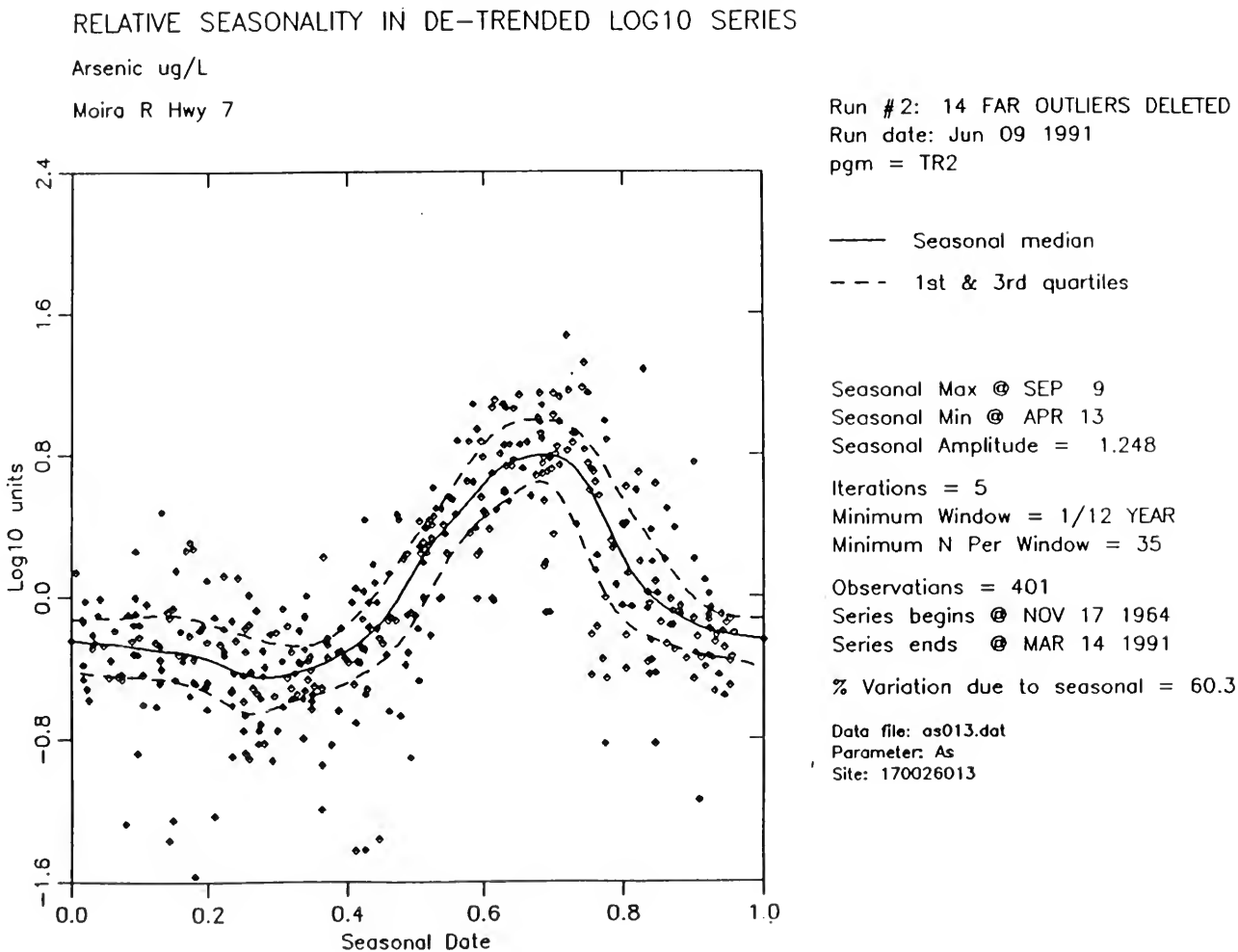


**Figure 9.c Seasonal trends in arsenic, Moira R.**
**Control parameters: $\widehat{m} = 70$ and $h = 6$**

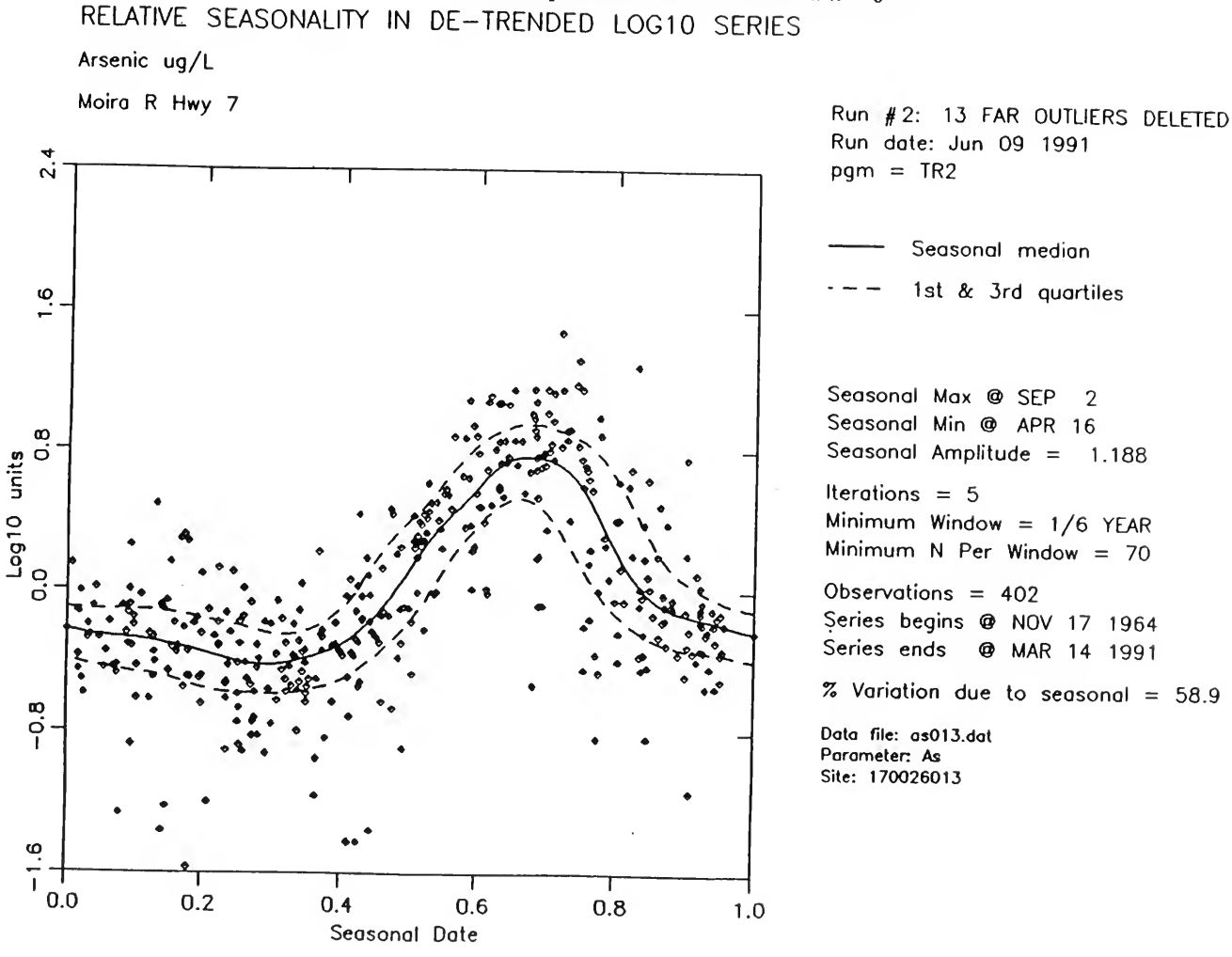


**Figure 9.d. Historical arsenic trends, Moira R.**

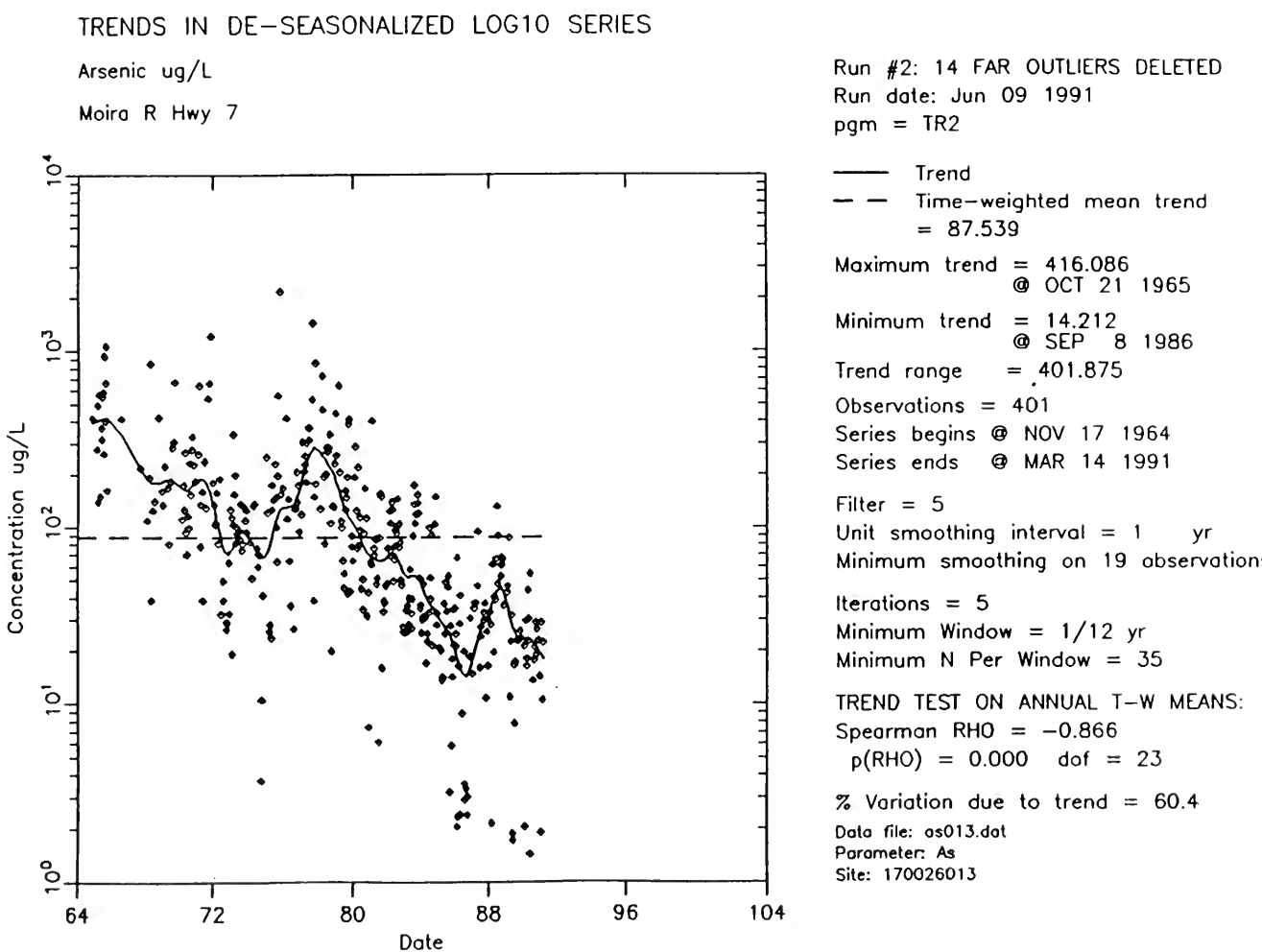


**Figure 10.1 Seasonal trends in total nitrates, Trent R.**

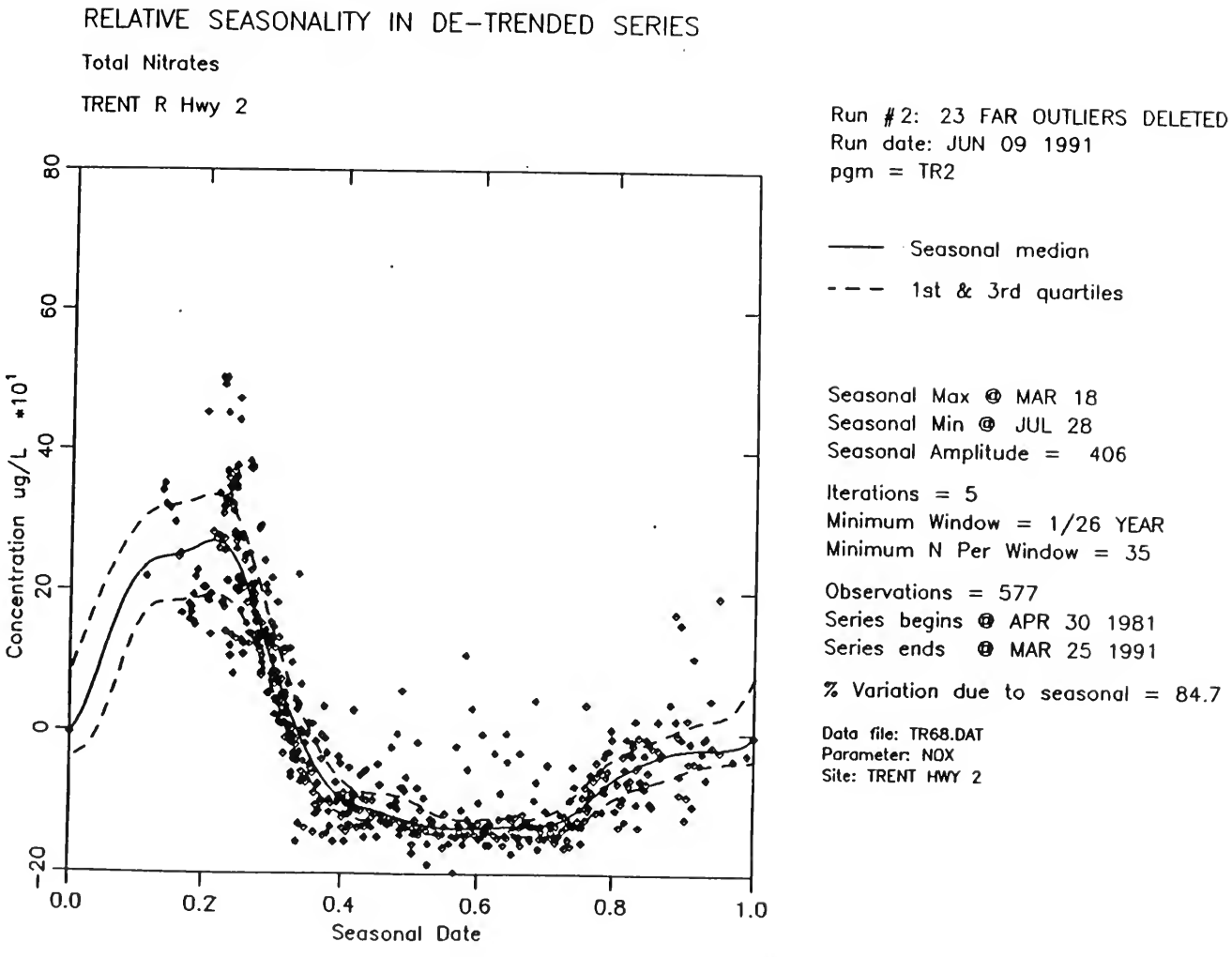


**Figure 10.2 Seasonal trends in nitrates, Plastic L inflows.**

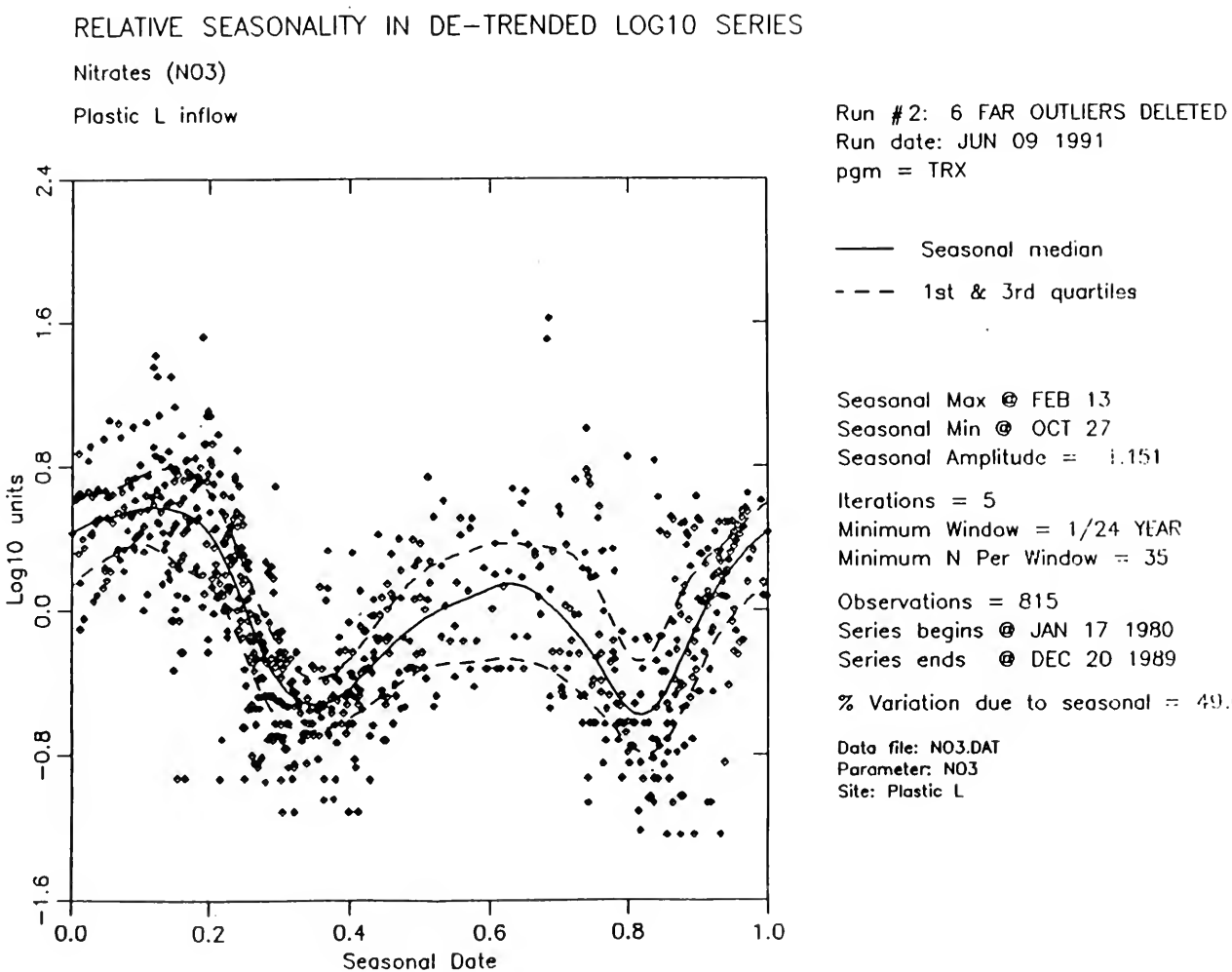


**Figure 11.1 Chloride trend plot, Rouge R, Run #1.**

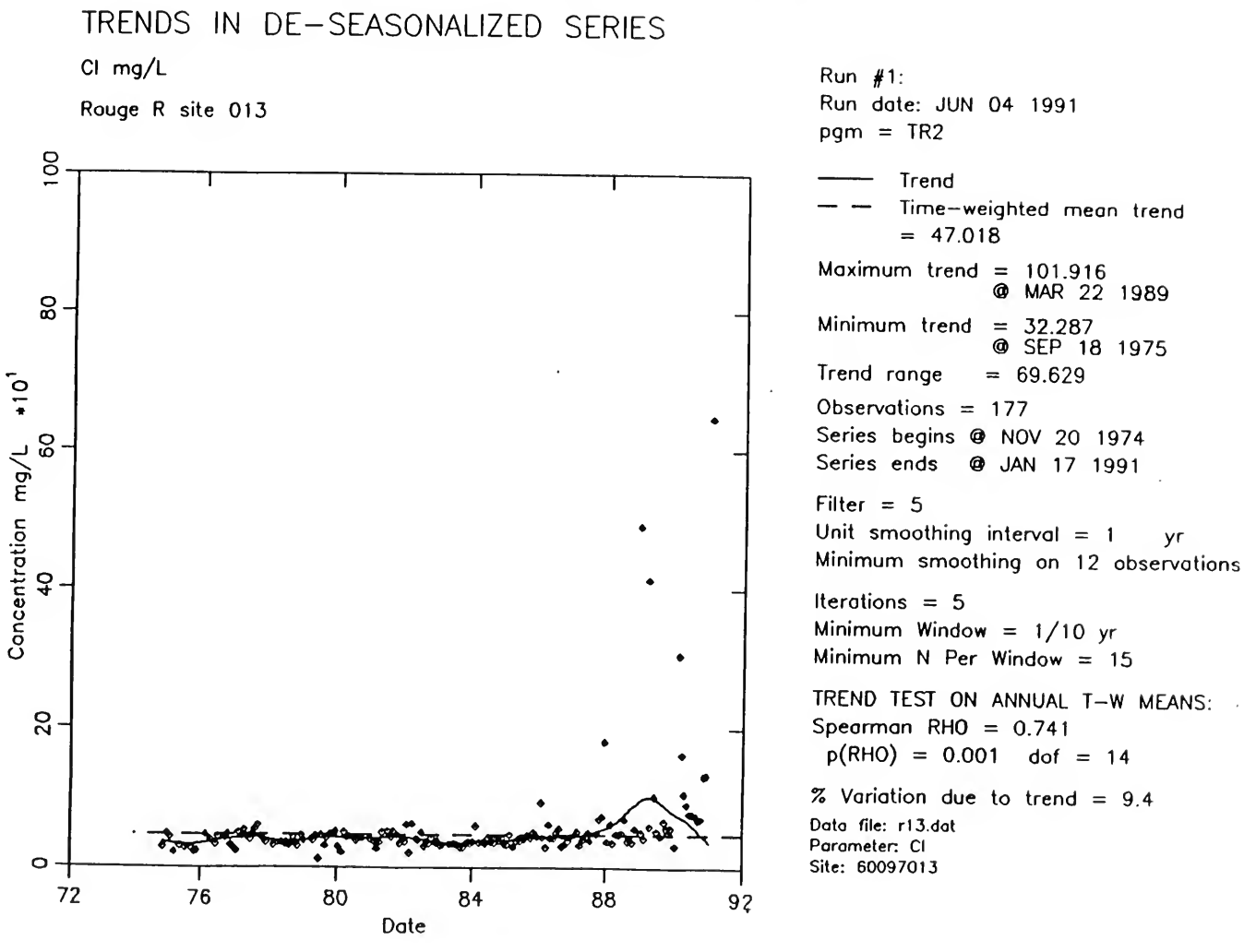


**Figure 11.2 Chloride seasonal plot, Rouge R, Run #1.**

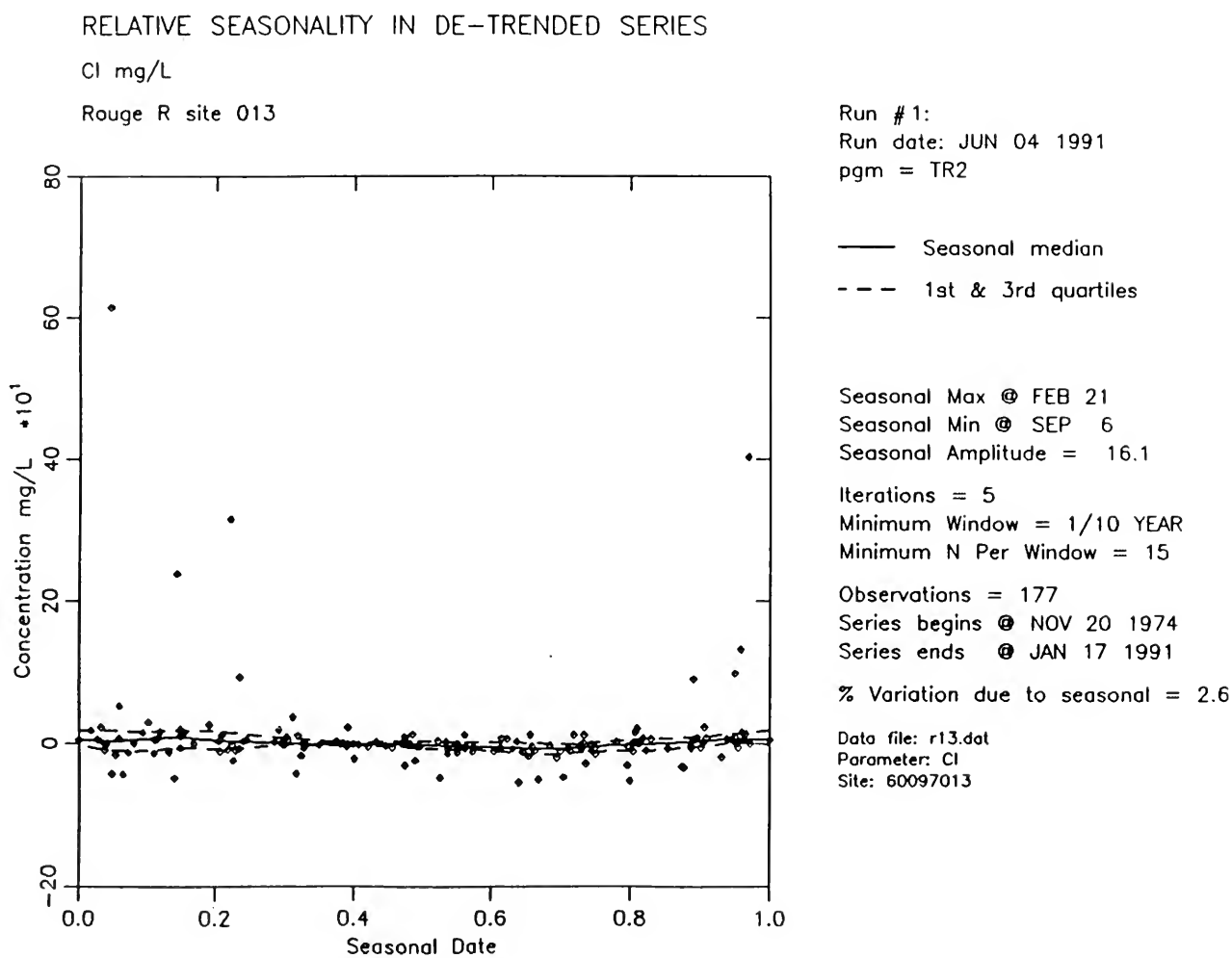


**Figure 11.3 Chloride trend plot, Rouge R, Run #2.**

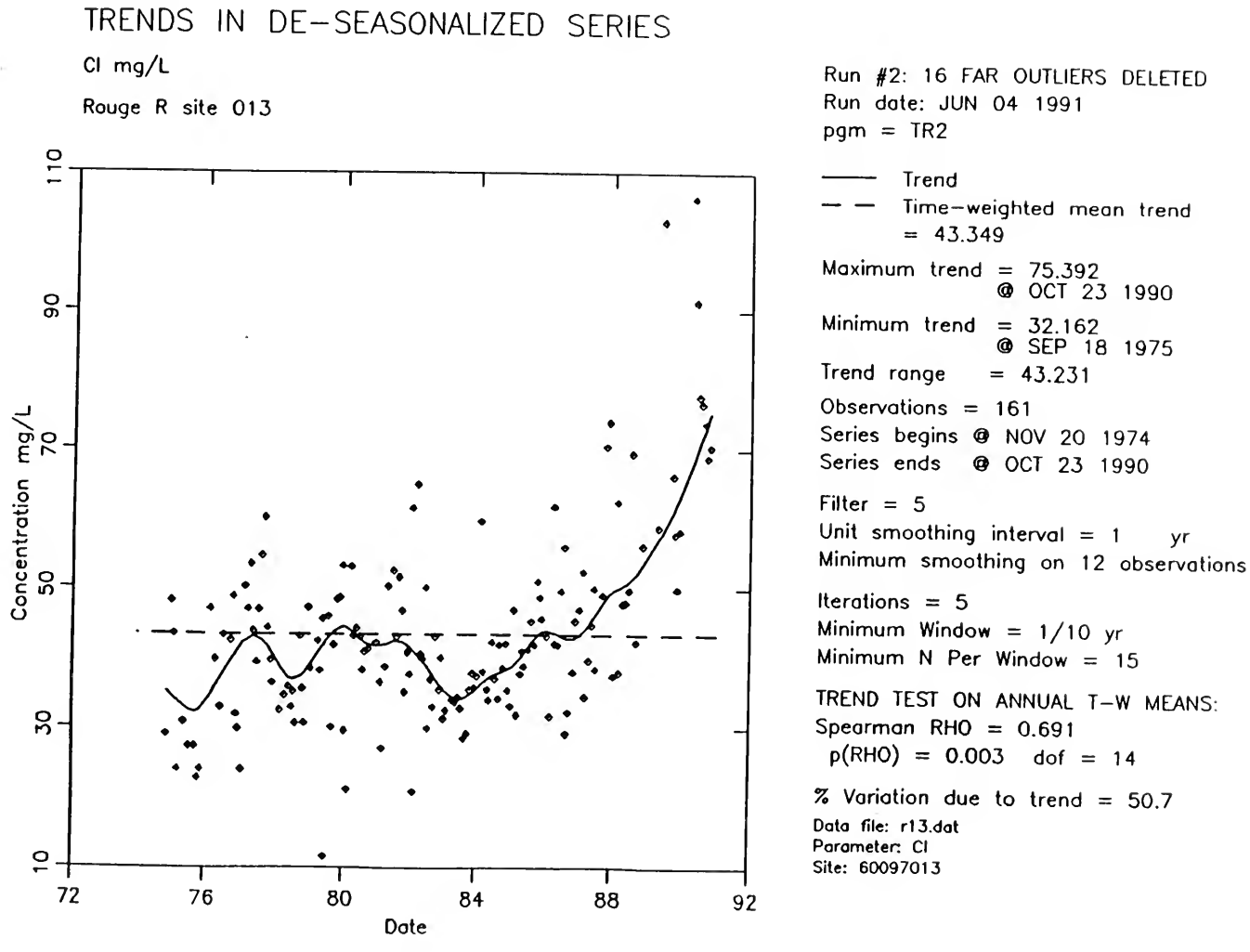


**Figure 11.4 Chloride seasonal plot, Rouge R, Run #2.**

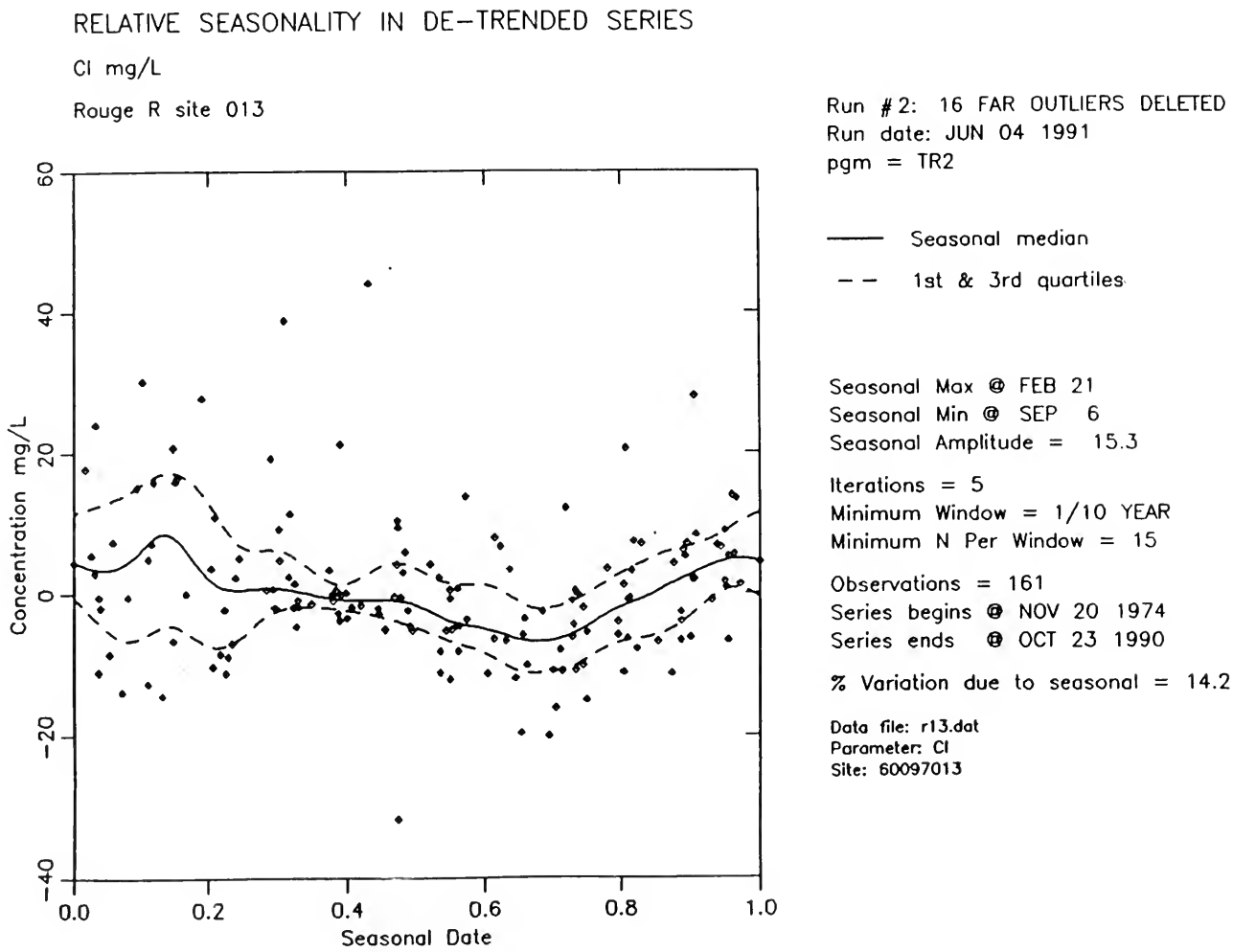
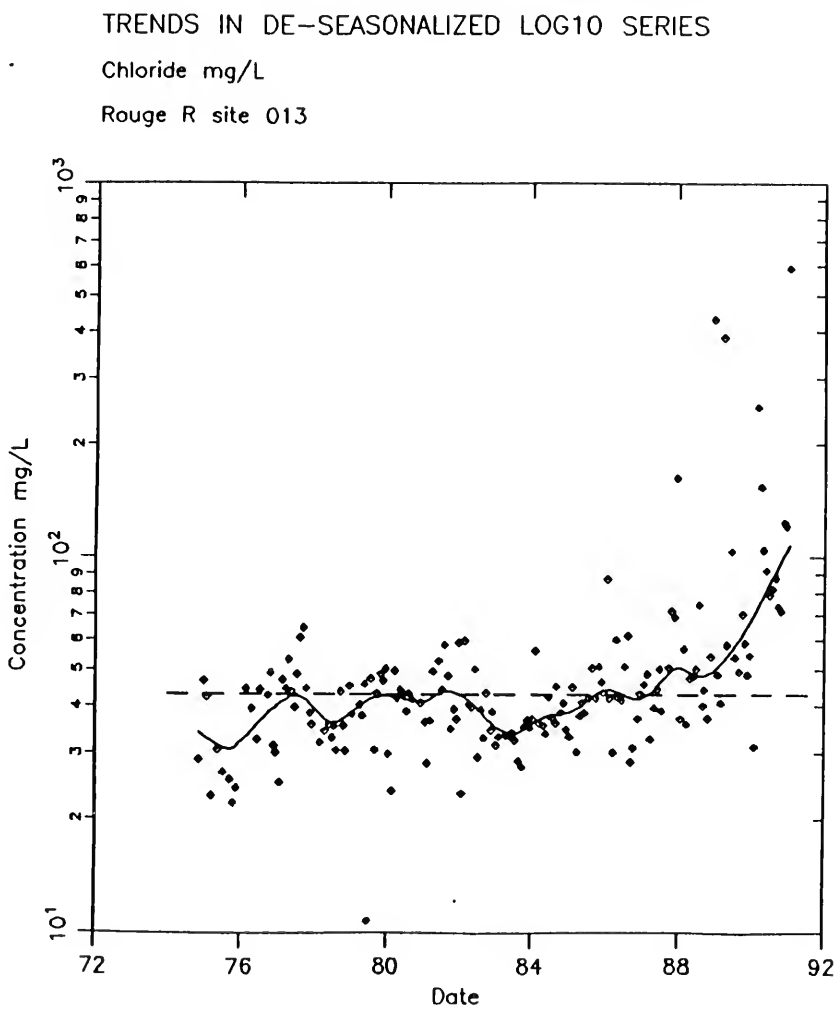


**Figure 11.5 $Log_{10}$ Chloride trend plot, Rouge R, Run #1.**




Run #1:
Run date: JUN 11 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 42.883

Maximum trend = 107.554
@ JAN 17 1991

Minimum trend = 30.522
@ SEP 18 1975

Trend range = 77.032

Observations = 177
Series begins @ NOV 20 1974
Series ends @ JAN 17 1991

Filter = 5
Unit smoothing interval = 1 yr
Minimum smoothing on 12 observations

Iterations = 5
Minimum Window = 1/10 yr
Minimum N Per Window = 15

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = 0.765
p(RHO) = 0.001 dof = 14

% Variation due to trend = 41.5

Data file: r13.dat
Parameter: P
Site: 60097013

**Figure 11.6 $Log_{10}$ Chloride seasonal plot, Rouge R, Run #1.**

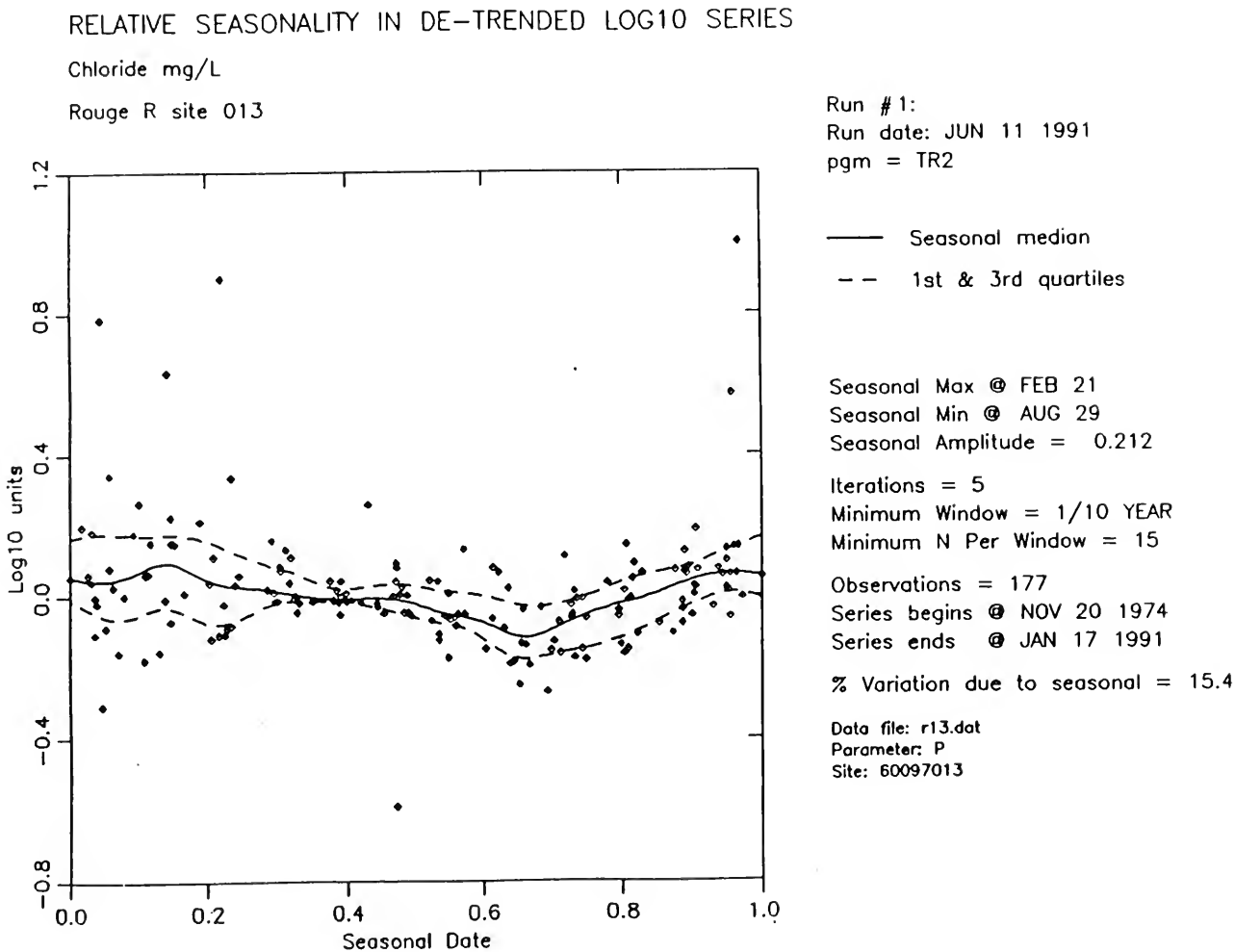


**Figure 11.7** $\text{Log}_{10}$ **Chloride trend plot, Rouge R, Run #2.**

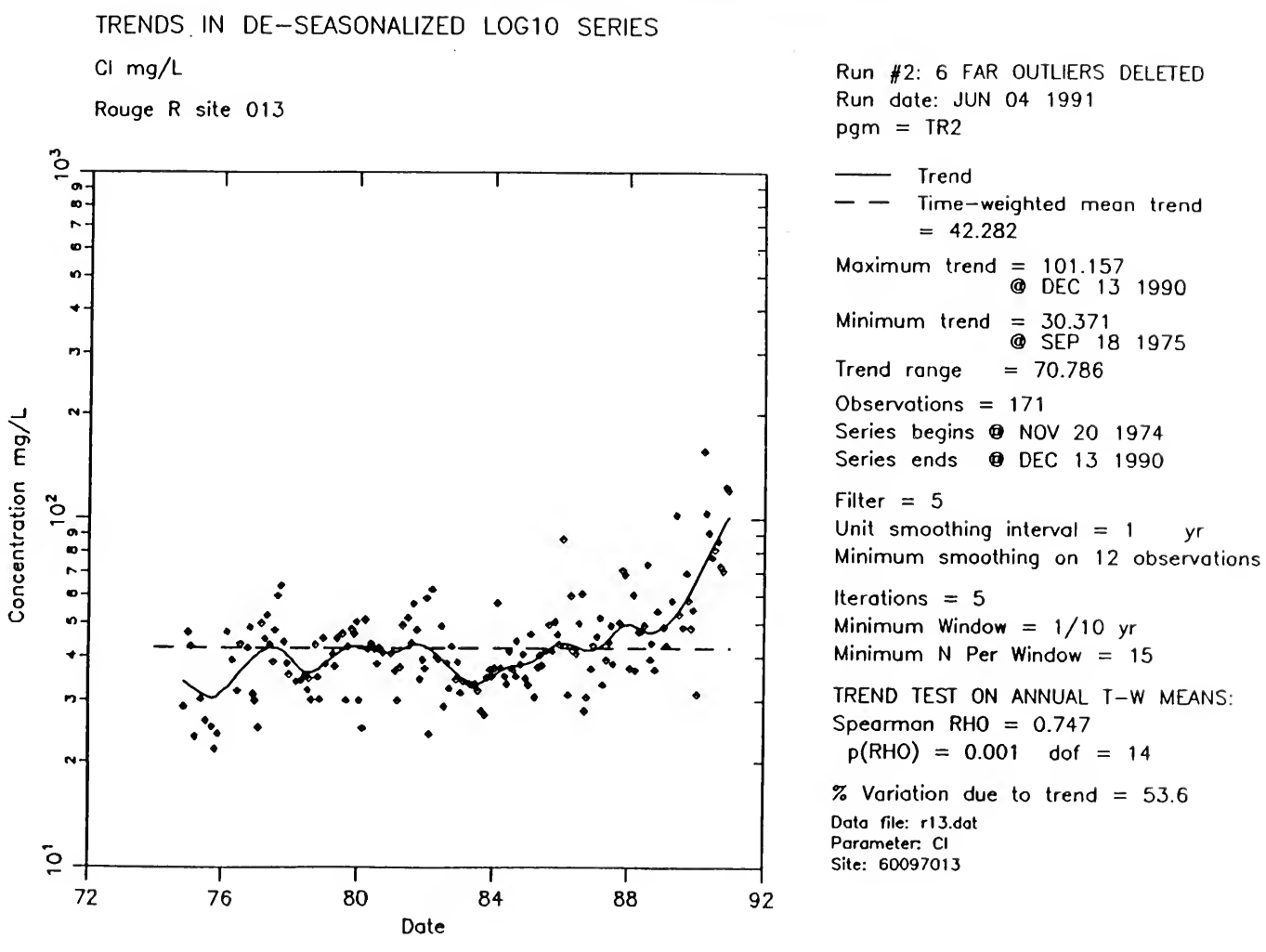


**Figure 11.8 $Log_{10}$ Chloride seasonal plot, Rouge R, Run #2.**

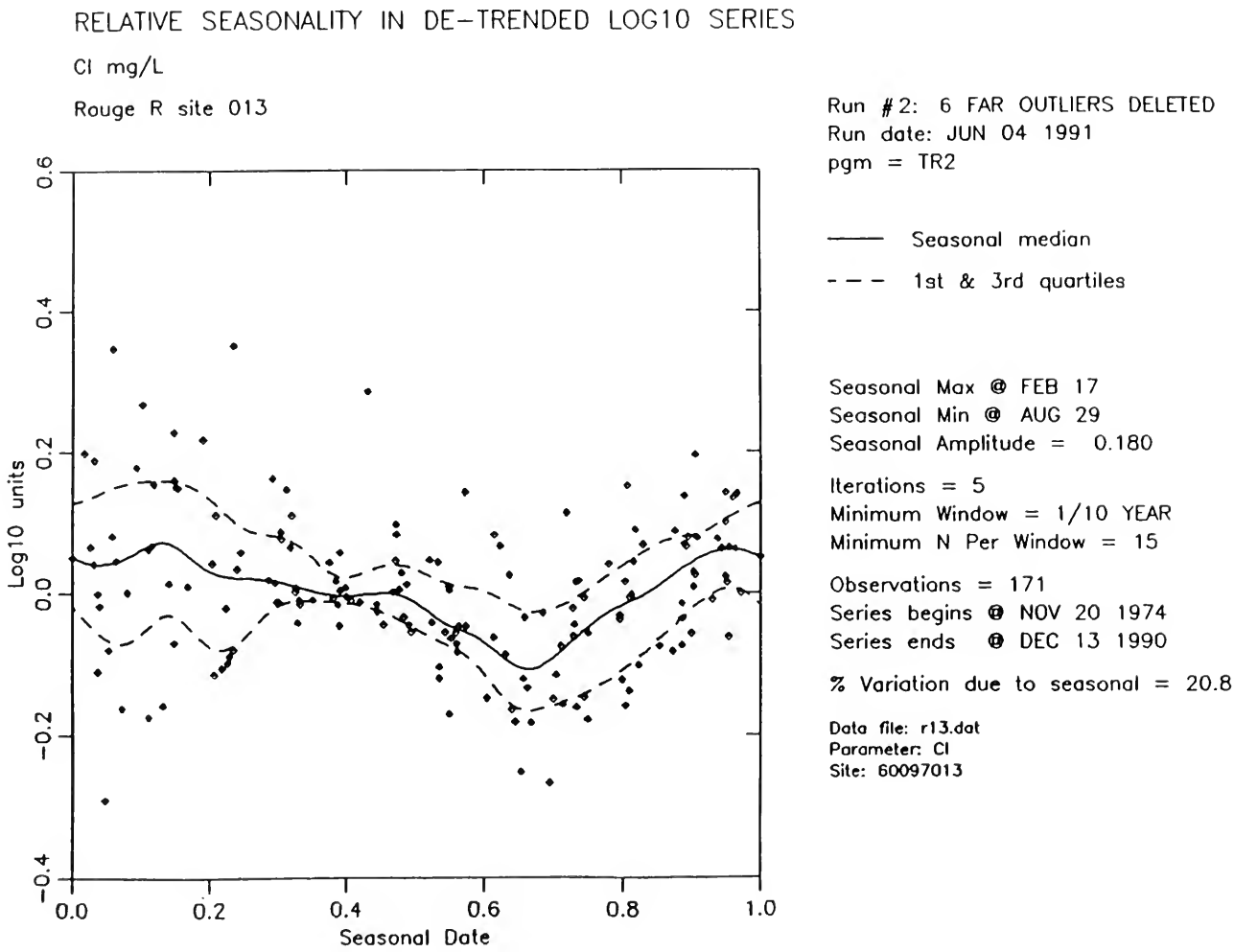


**Figure 12.1 Heart L total P trend plot, 1 iteration.**

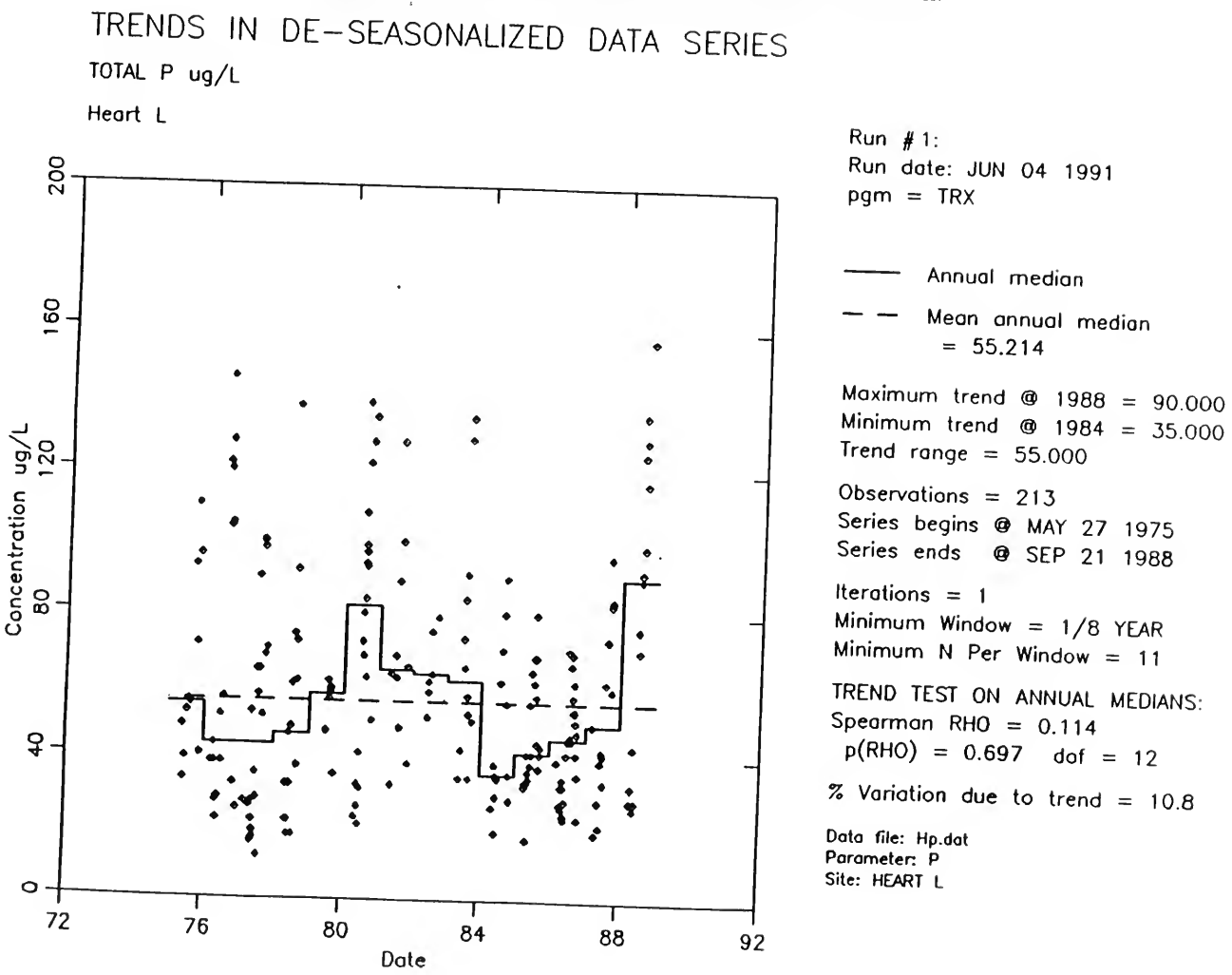


**Figure 12.2 Heart L total P trend plot, 2 iterations.**

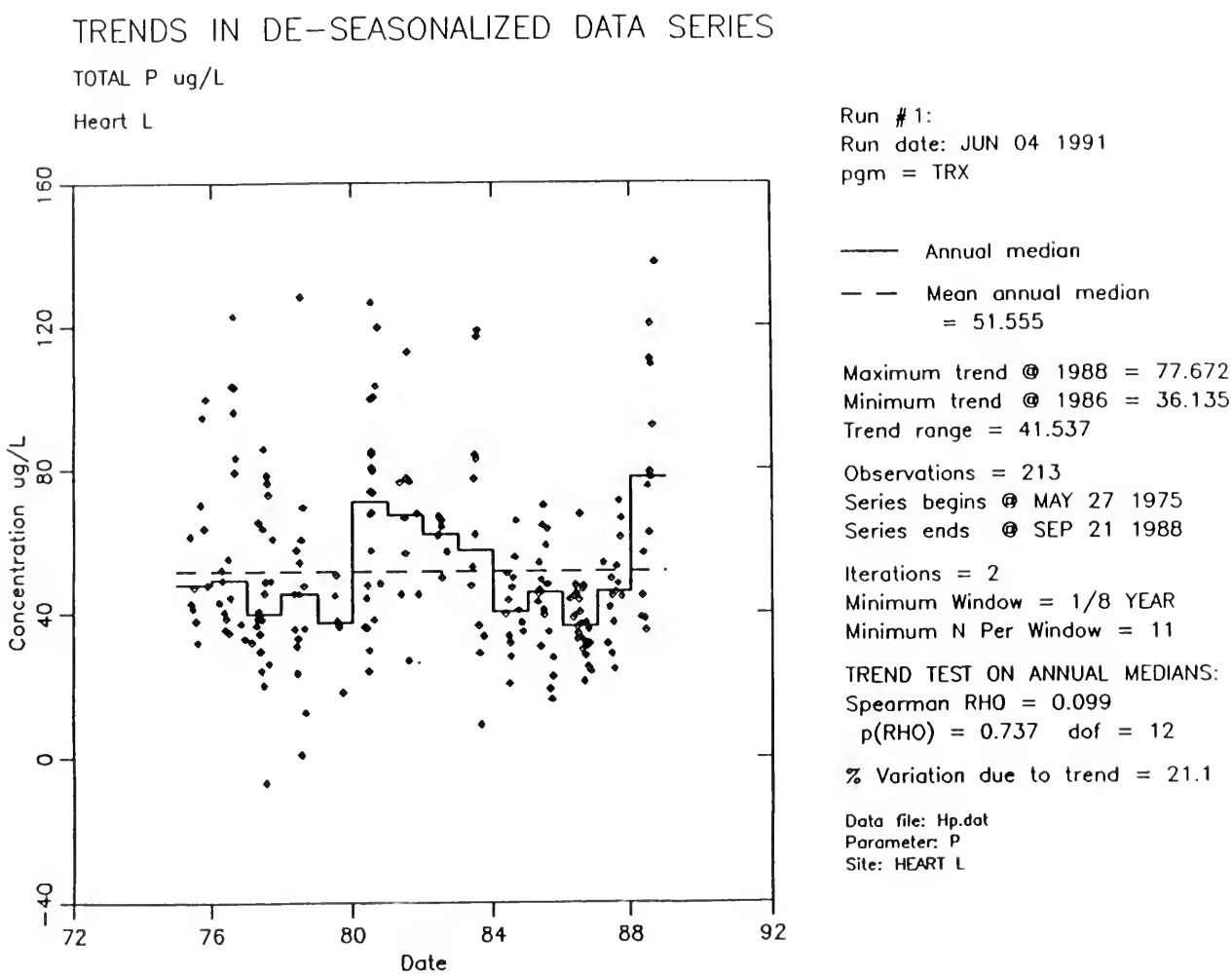


**Figure 12.3 Heart L total P trend plot, 3 iterations.**

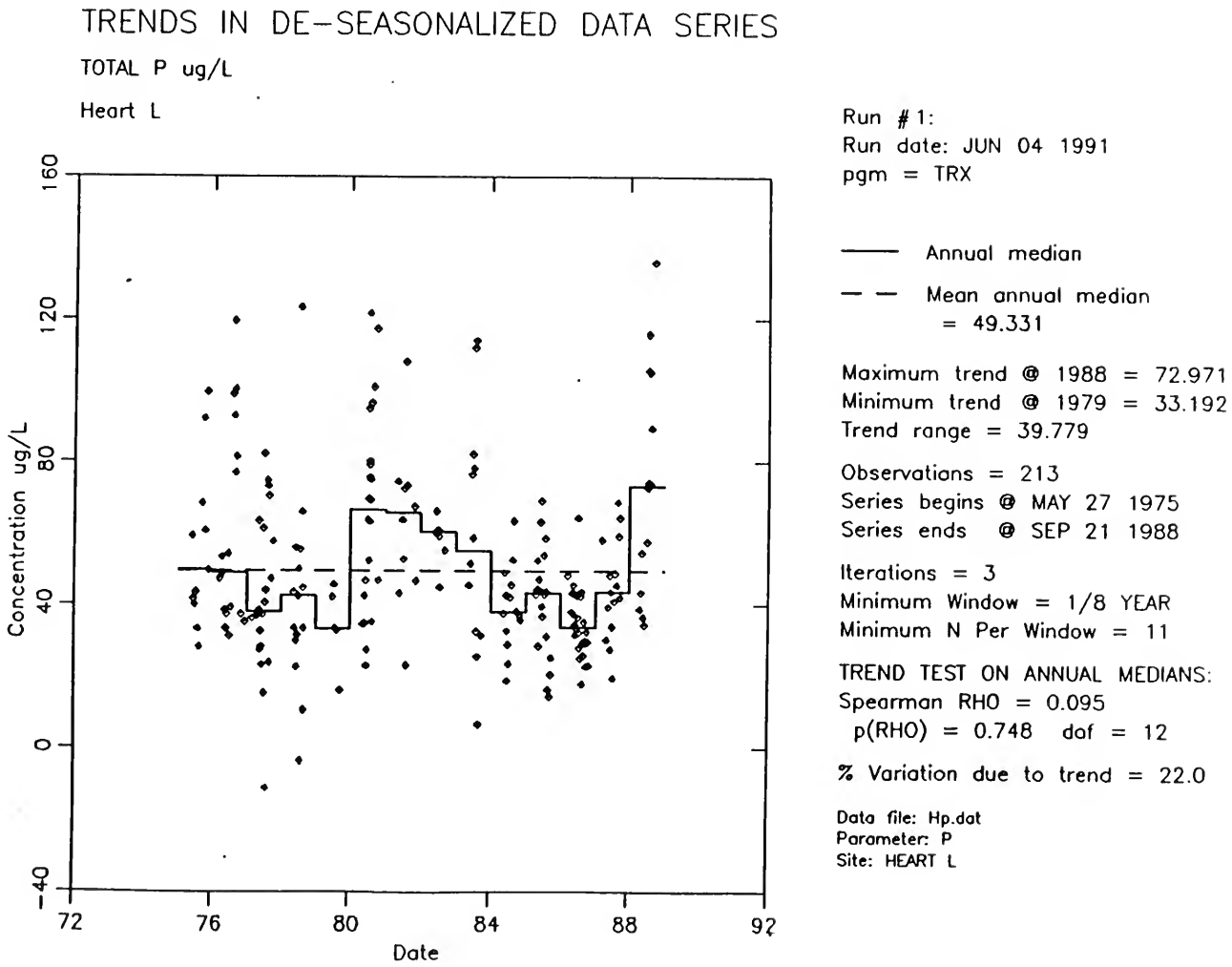


**Figure 12.4 Heart L total P trend plot, 5 iterations.**

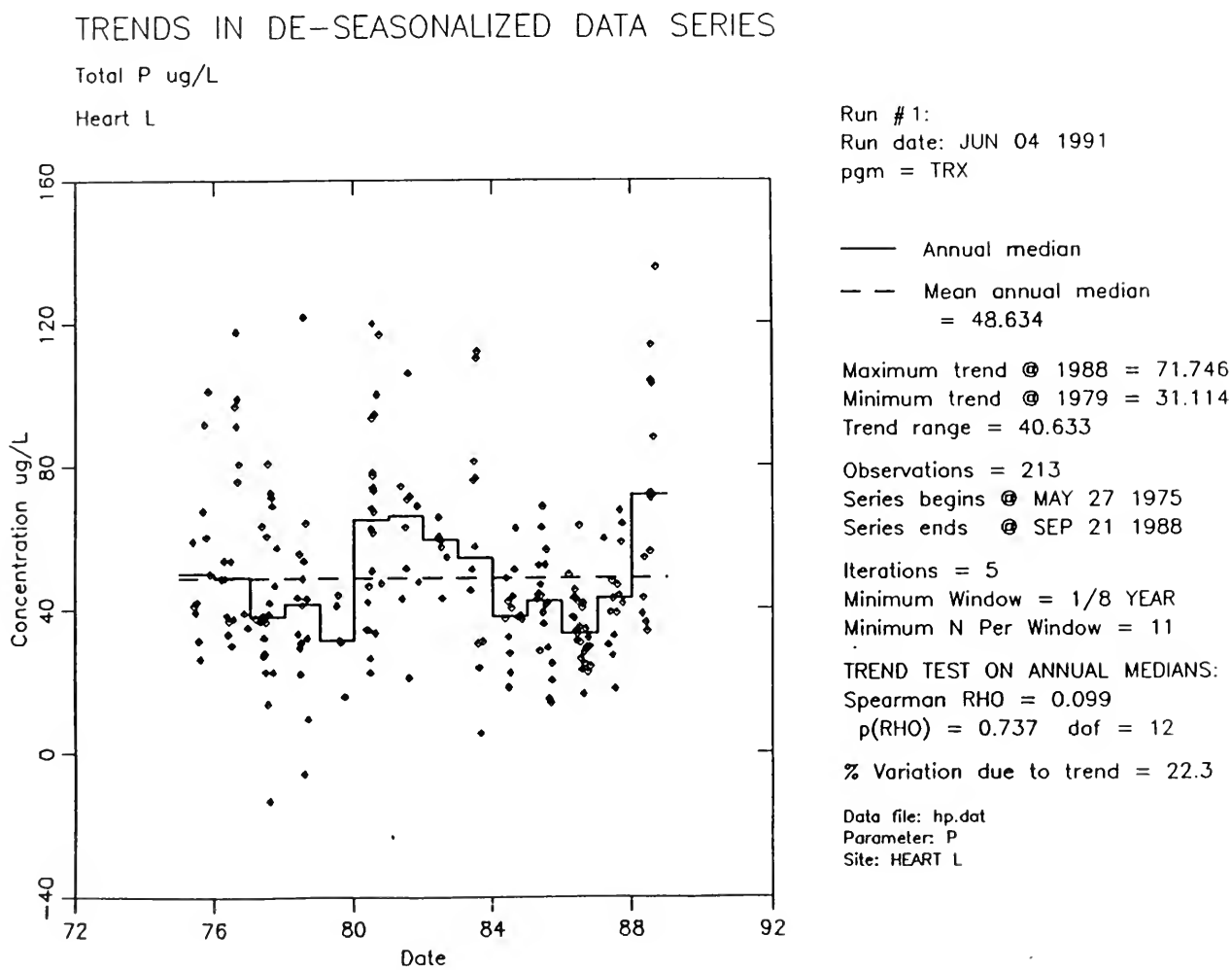


**Figure 12.5 Heart L total P trend plot, 10 iterations.**

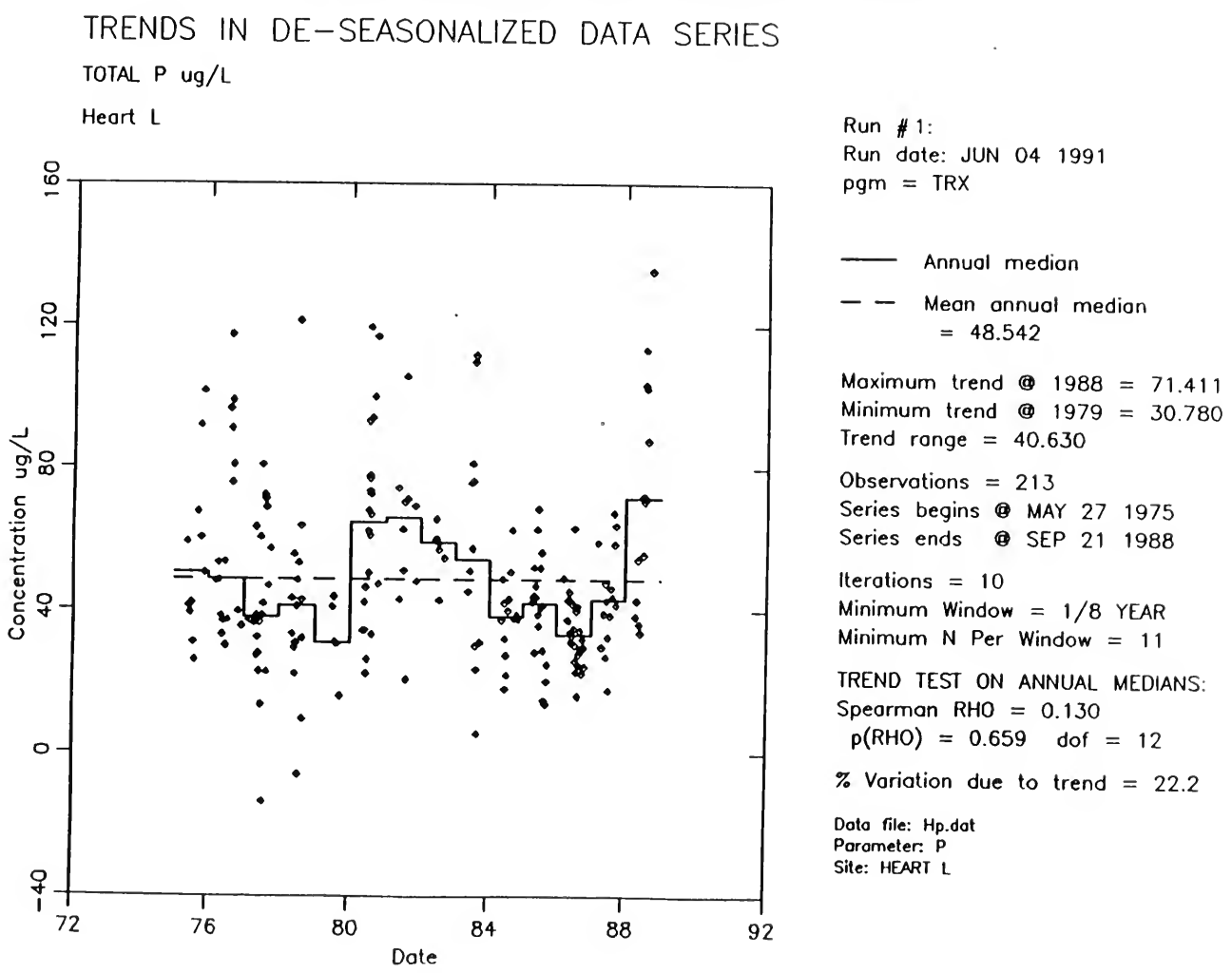


**Figure 12.6 Heart L total P seasonal plot, 1 iteration.**

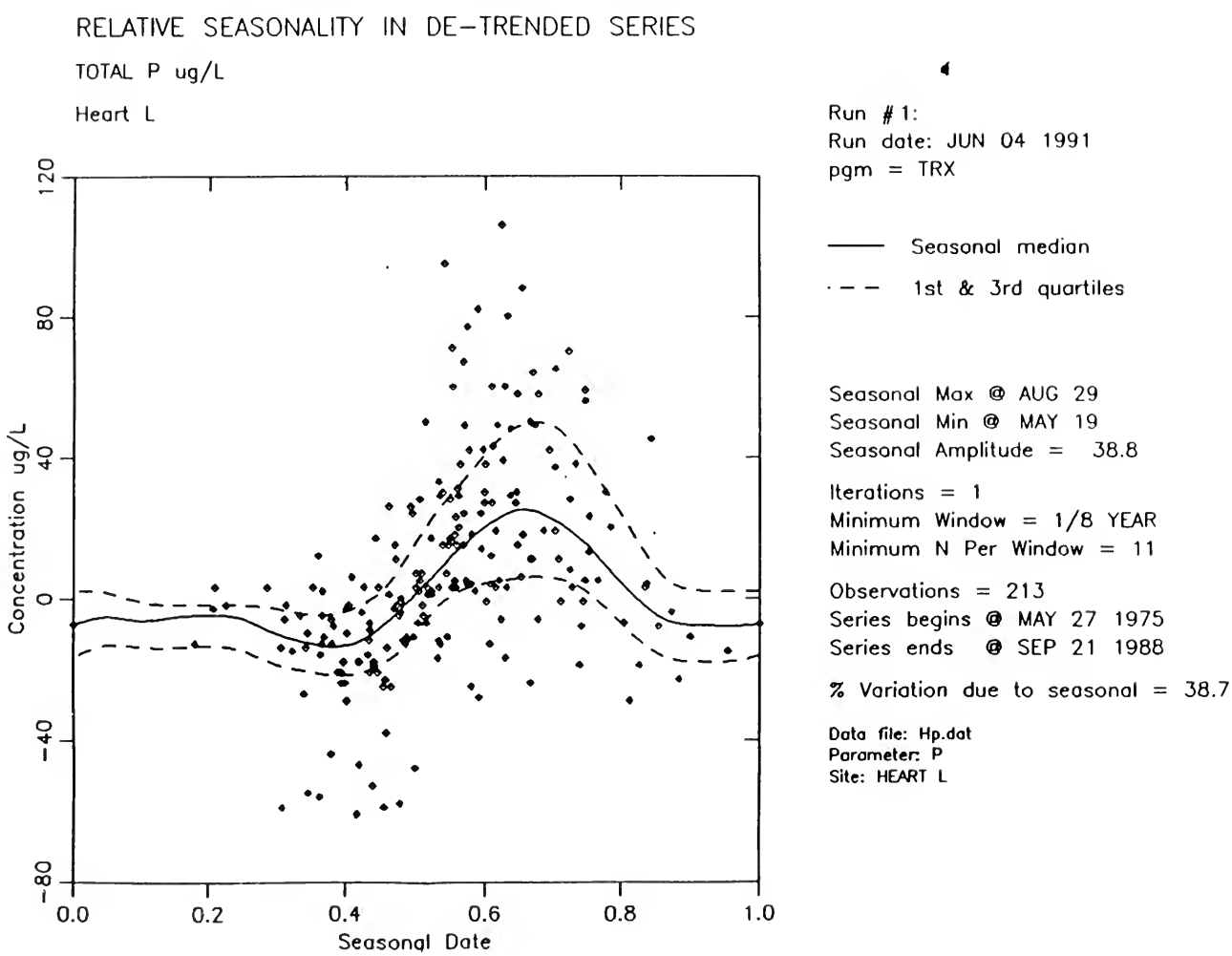


**Figure 12.7 Heart L total P seasonal plot, 2 iterations.**

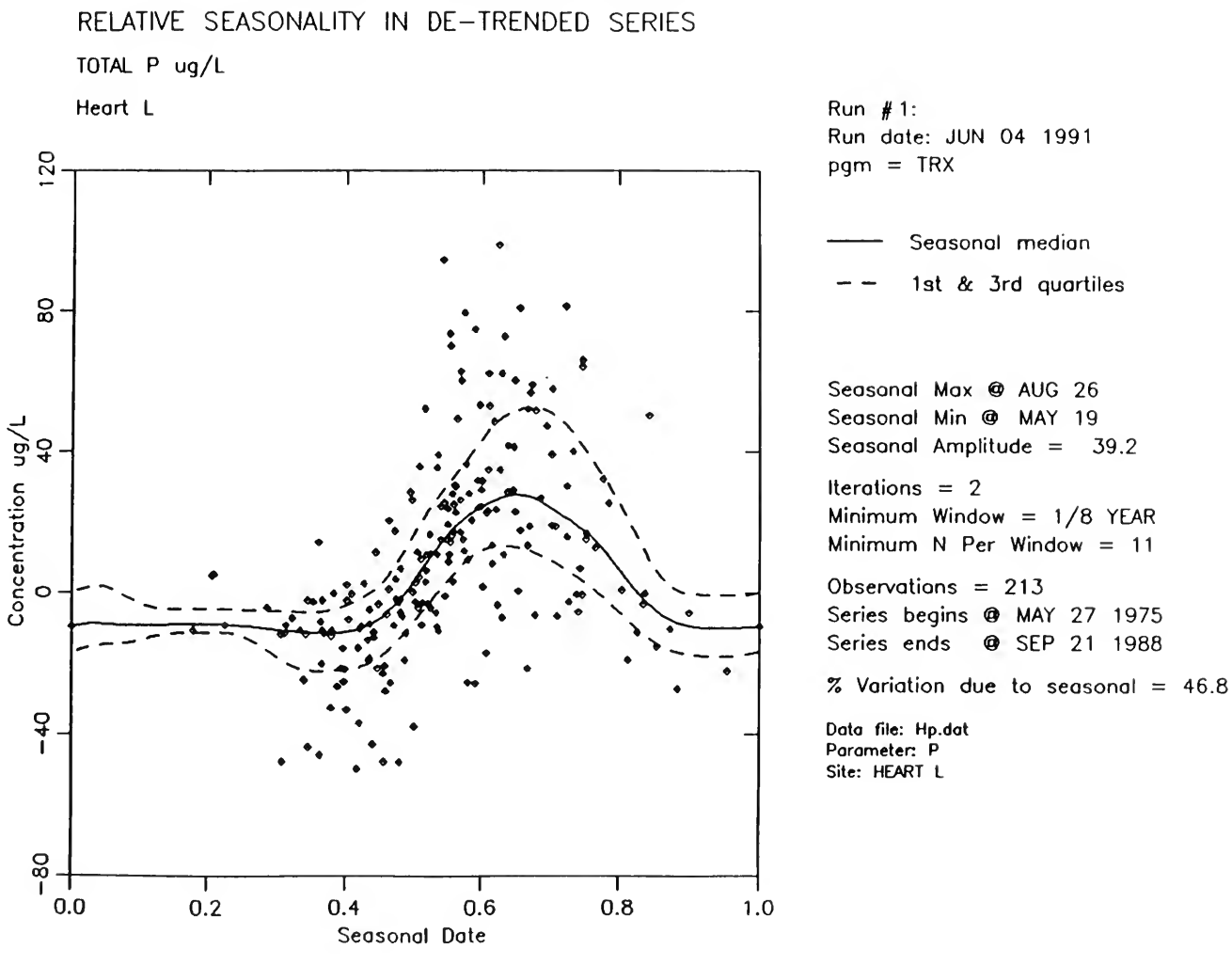


**Figure 12.8 Heart L total P seasonal plot, 10 iterations.**

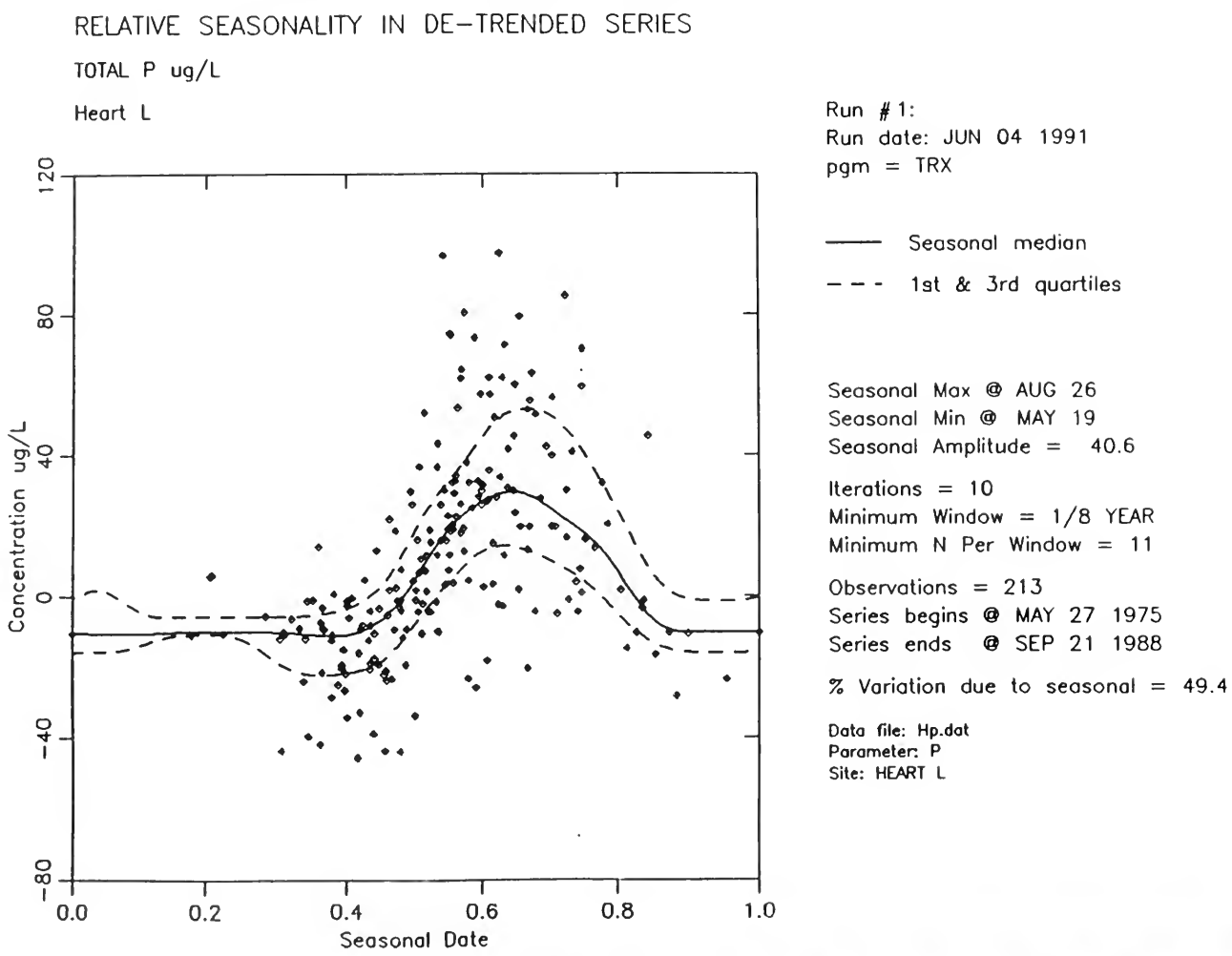


**Figure 13.1 Na series with dubious early records & 1990 singleton.**

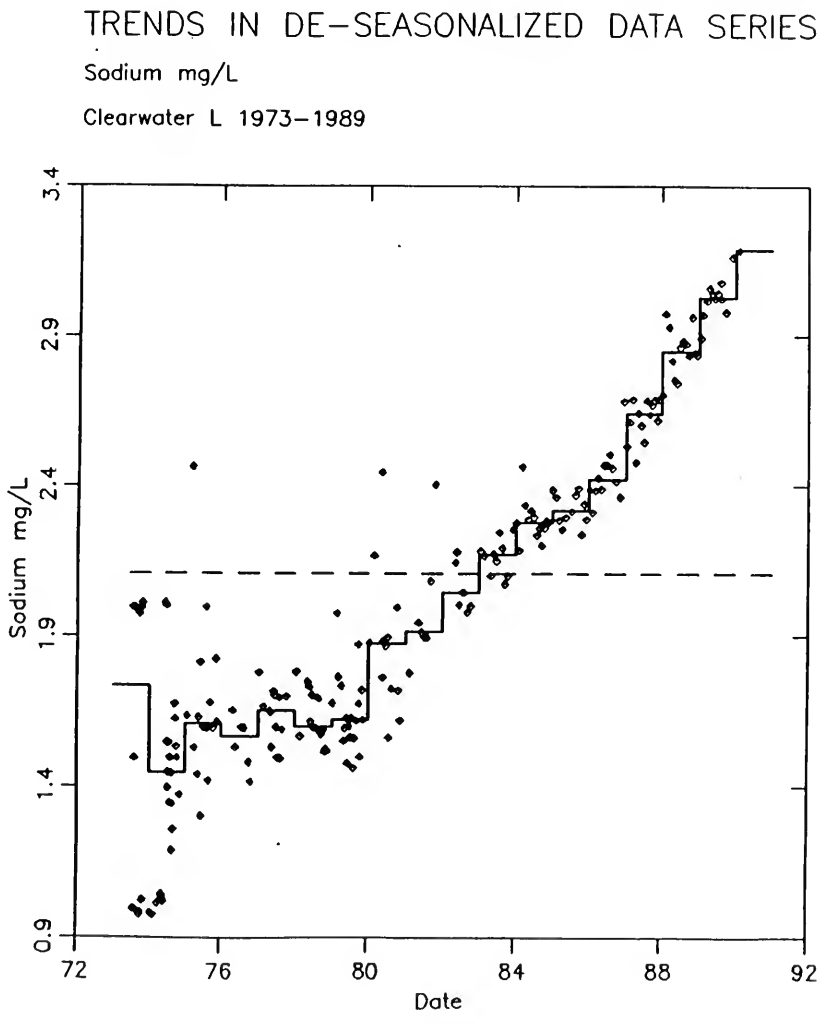
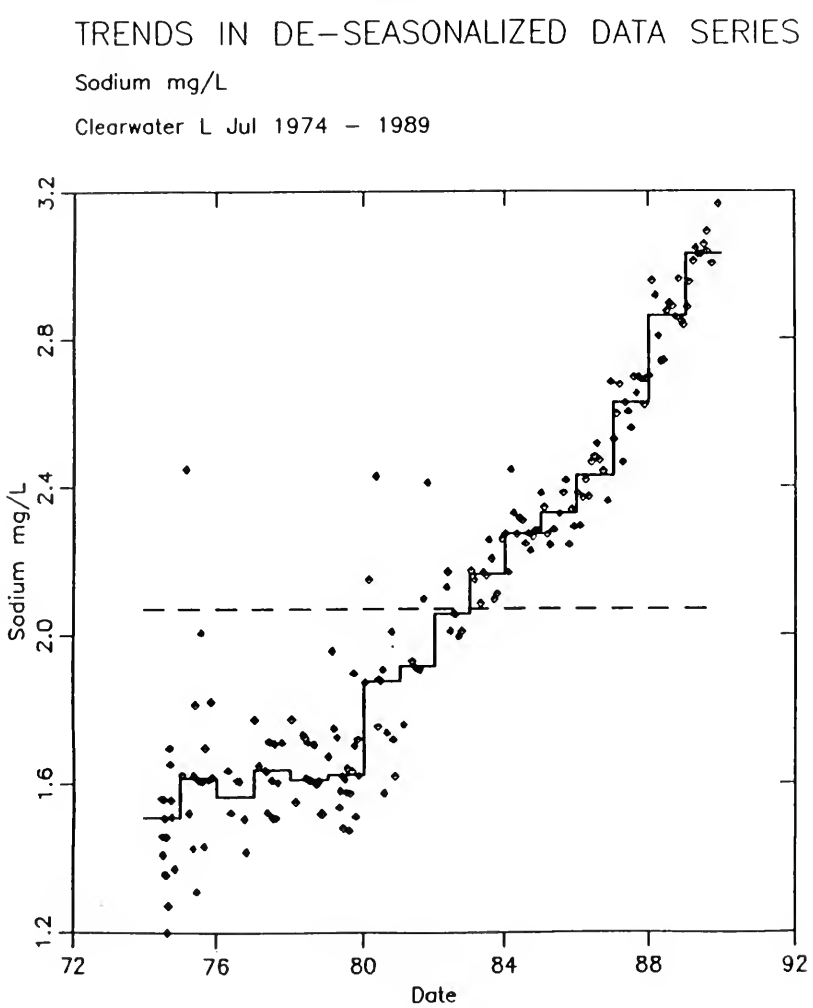


**Figure 13.2 Na series restricted to reliable period of record & the end of 1989.**




Run # 1:
Run date: MAY 23 1991
pgm = TRX

——— Annual median

— — Mean annual median
= 2.069

Maximum trend @ 1989 = 3.029
Minimum trend @ 1974 = 1.506
Trend range = 1.523

Observations = 197
Series begins @ JUL 3 1974
Series ends @ NOV 28 1989

Iterations = 5
Minimum Window = 1/8 YEAR
Minimum N Per Window = 13

TREND TEST ON ANNUAL MEDIANS:
Spearman RHO = 0.979
p(RHO) = 0.000 dof = 14

% Variation due to trend = 93.0

Data file: Na.dat
Parameter: NA
Site: Clearwater L

**Figure 14.1 Fe series with singleton data at 1973 and 1990.**

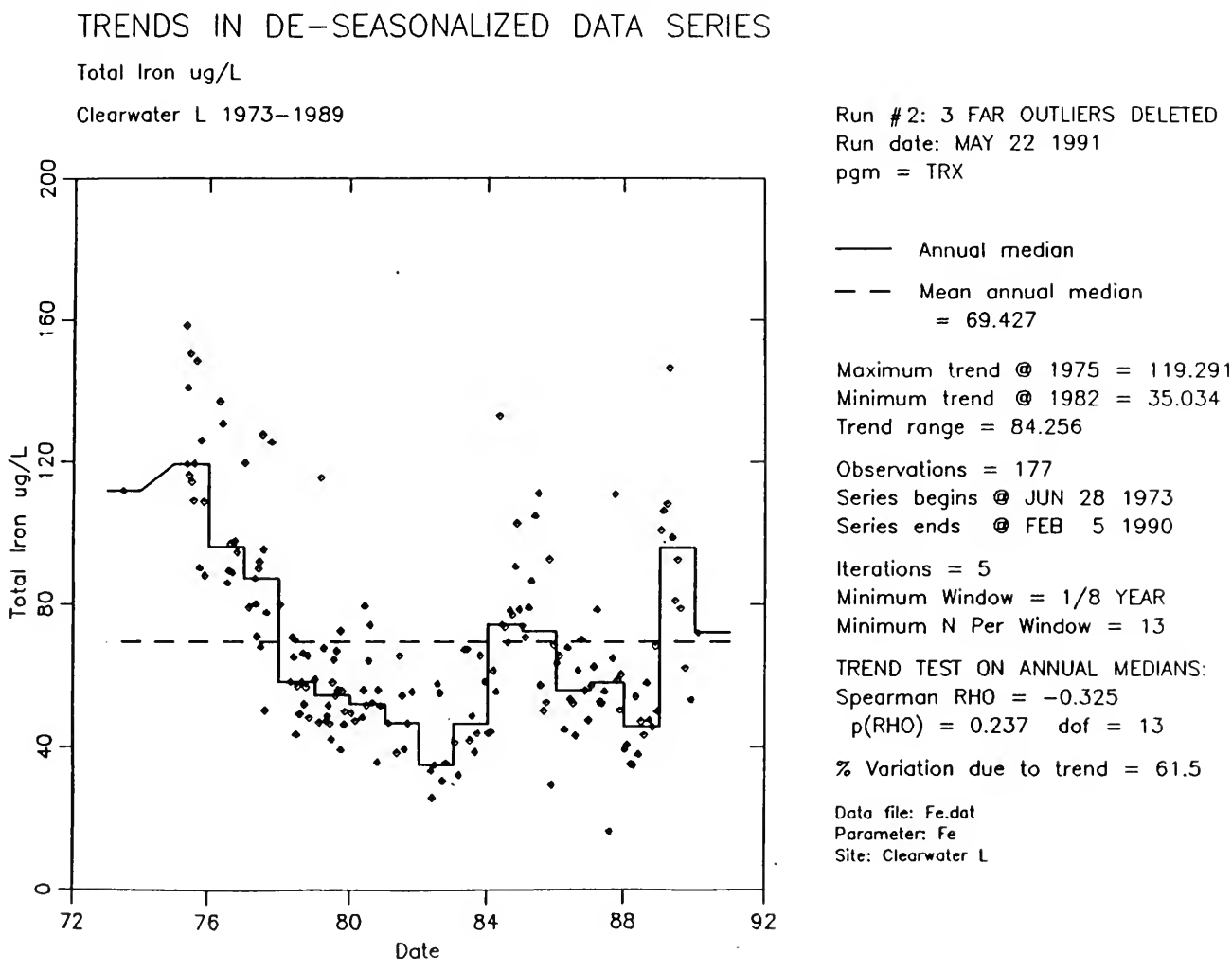
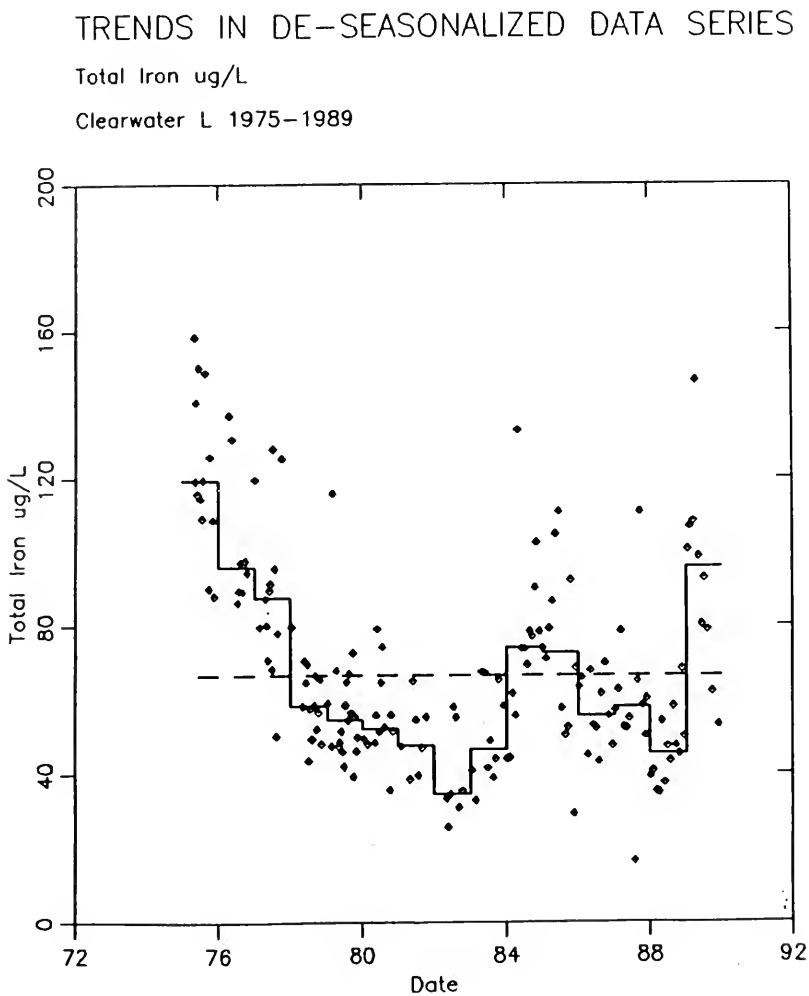


**Figure 14.2 Fe series bracketed to 1975-1989.**




Run #2: 3 FAR OUTLIERS DELETED
Run date: MAY 22 1991
pgm = TRX

——— Annual median

— — Mean annual median
= 66.397

Maximum trend @ 1975 = 119.193
Minimum trend @ 1982 = 34.445
Trend range = 84.747

Observations = 175
Series begins @ MAY 13 1975
Series ends @ NOV 28 1989

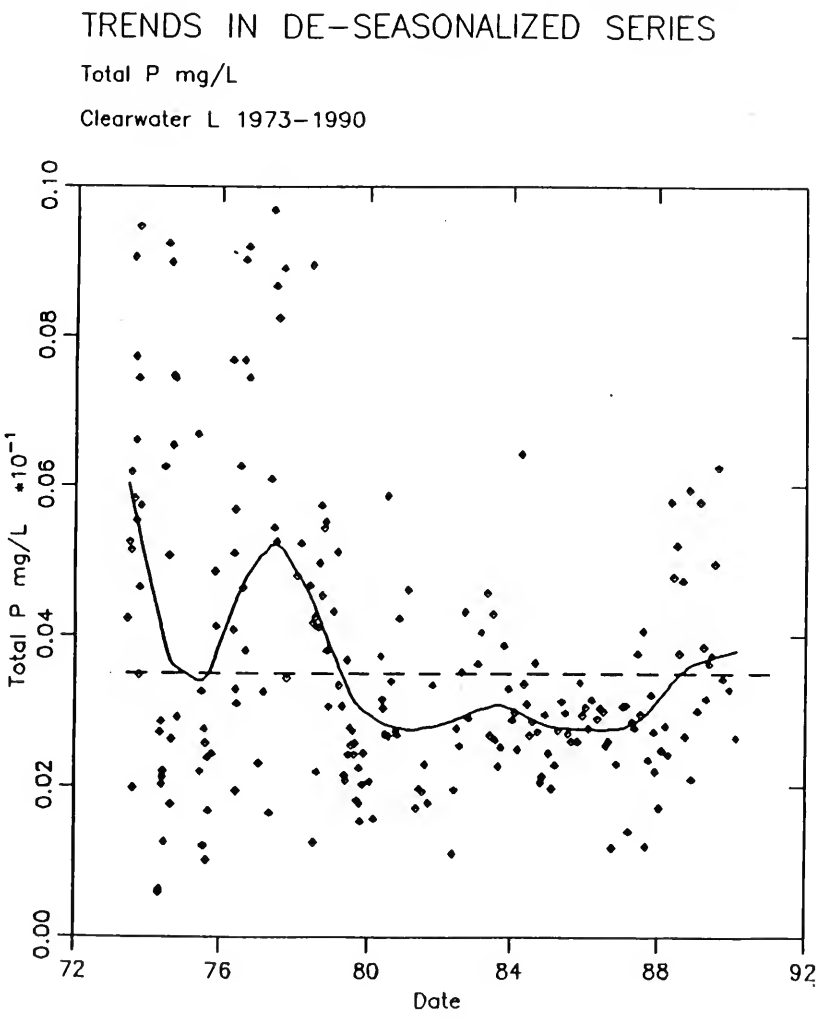
Iterations = 5
Minimum Window = 1/8 YEAR
Minimum N Per Window = 13

TREND TEST ON ANNUAL MEDIANS:
Spearman RHO = −0.325
p(RHO) = 0.237 dof = 13

% Variation due to trend = 61.3

Data file: Fe.dat
Parameter: Fe
Site: Clearwater L

**Figure 15.1 Total P series in mg/L units.**




Run #3: 6 FAR OUTLIERS DELETED
Run date: MAY 22 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 0.003

Maximum trend = 0.006
@ JUN 28 1973

Minimum trend = 0.003
@ FEB 18 1981

Trend range = 0.003

Observations = 212
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 5
Unit smoothing interval = 2 yr
Minimum smoothing on 28 observations

Iterations = 3
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.575
p(RHO) = 0.016 dof = 15

% Variation due to trend = 30.7

Data file: P.DAT
Parameter: P
Site: Clearwater L

**Figure 15.2 Total P series in $\mu$g/L units.**

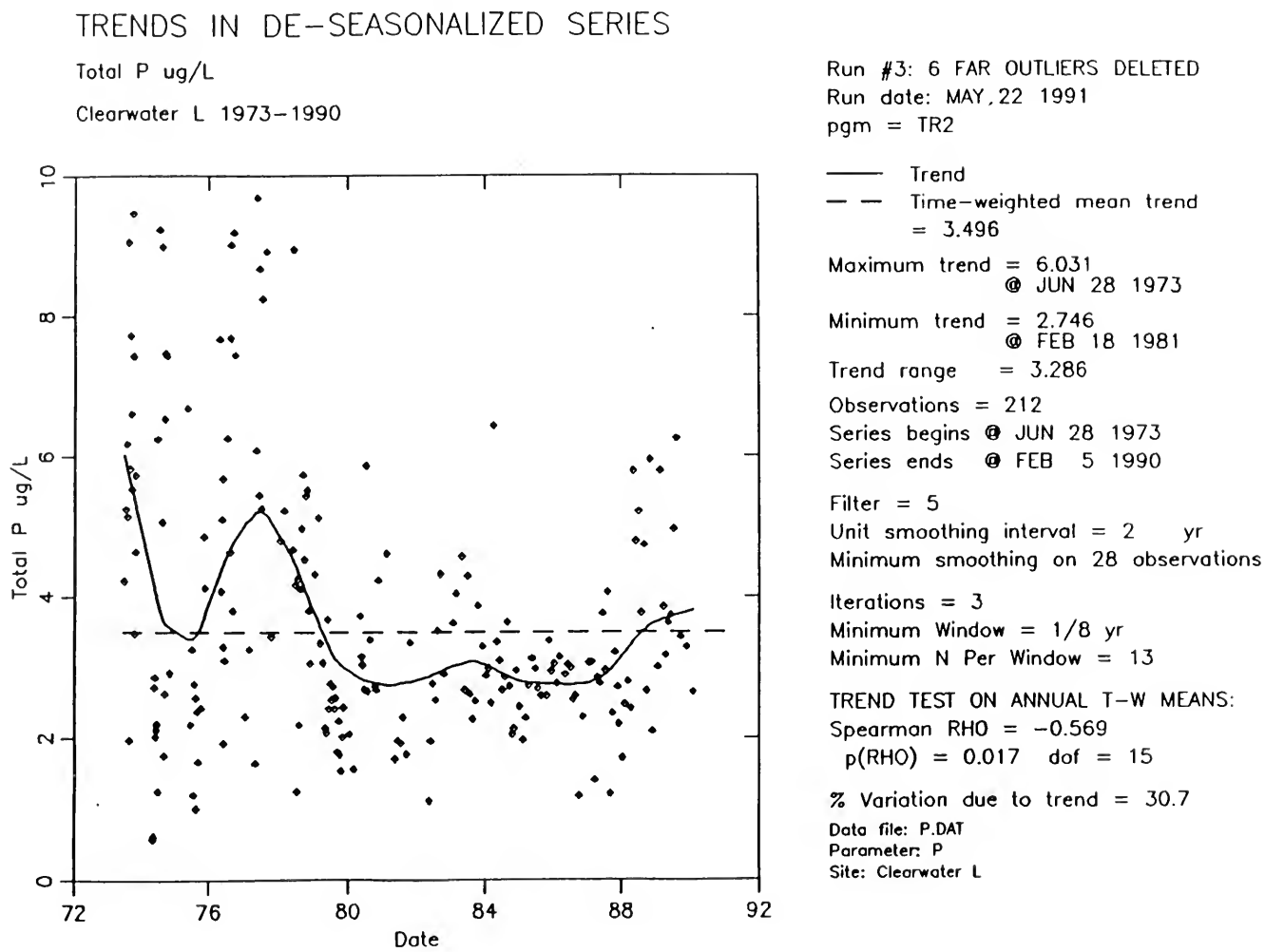
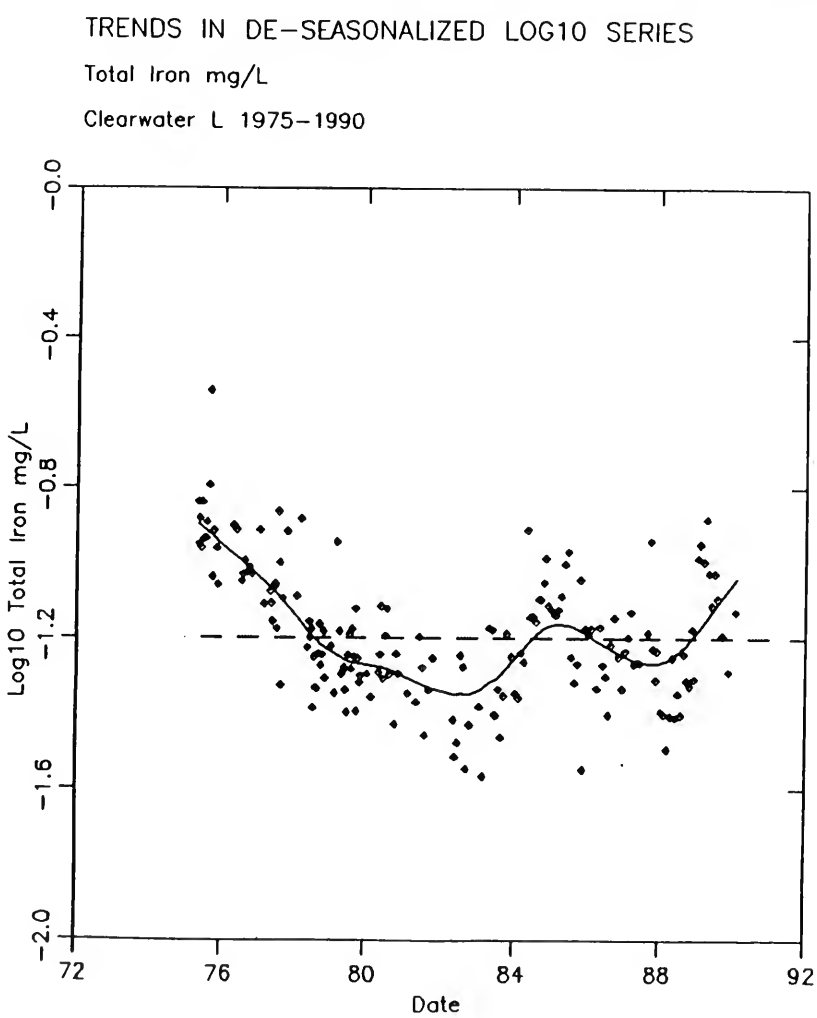


**Figure 15.3 Fe series in $\log_{10}$ mg/L units.**




Run #2: 2 FAR OUTLIERS DELETED
Run date: MAY 22 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 0.063

Maximum trend = 0.126
@ MAY 13 1975

Minimum trend = 0.045
@ OCT 18 1982

Trend range = 0.082

Observations = 177
Series begins @ MAY 13 1975
Series ends @ FEB 5 1990

Filter = 5
Unit smoothing interval = 2 yr
Minimum smoothing on 27 observations

Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = -0.304
p(RHO) = 0.271 dof = 13

% Variation due to trend = 55.2

Data file: Fe.DAT
Parameter: Fe
Site: Clearwater L

**Figure 15.4 Fe series in $\log_{10}$ $\mu$g/L units.**

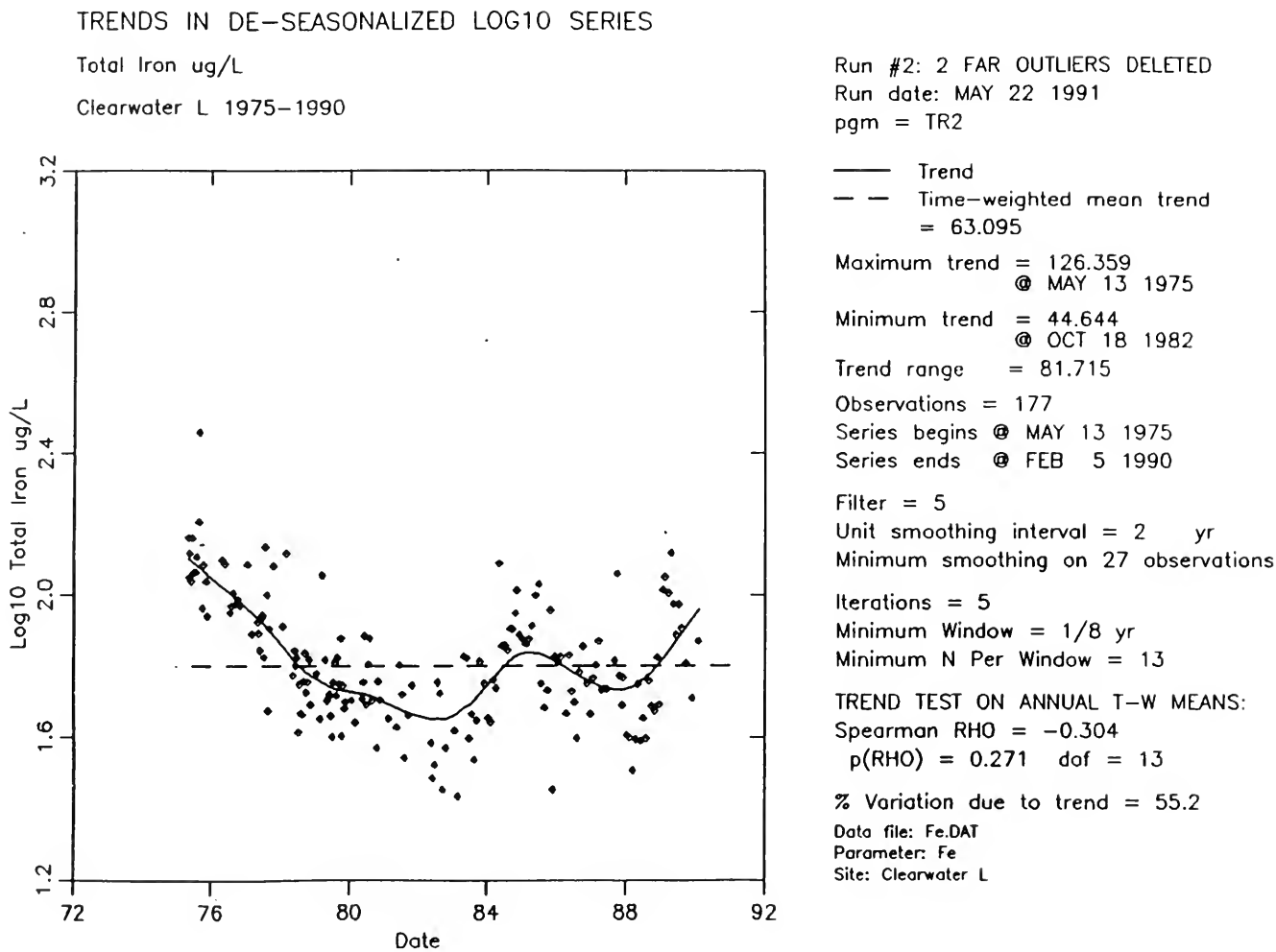
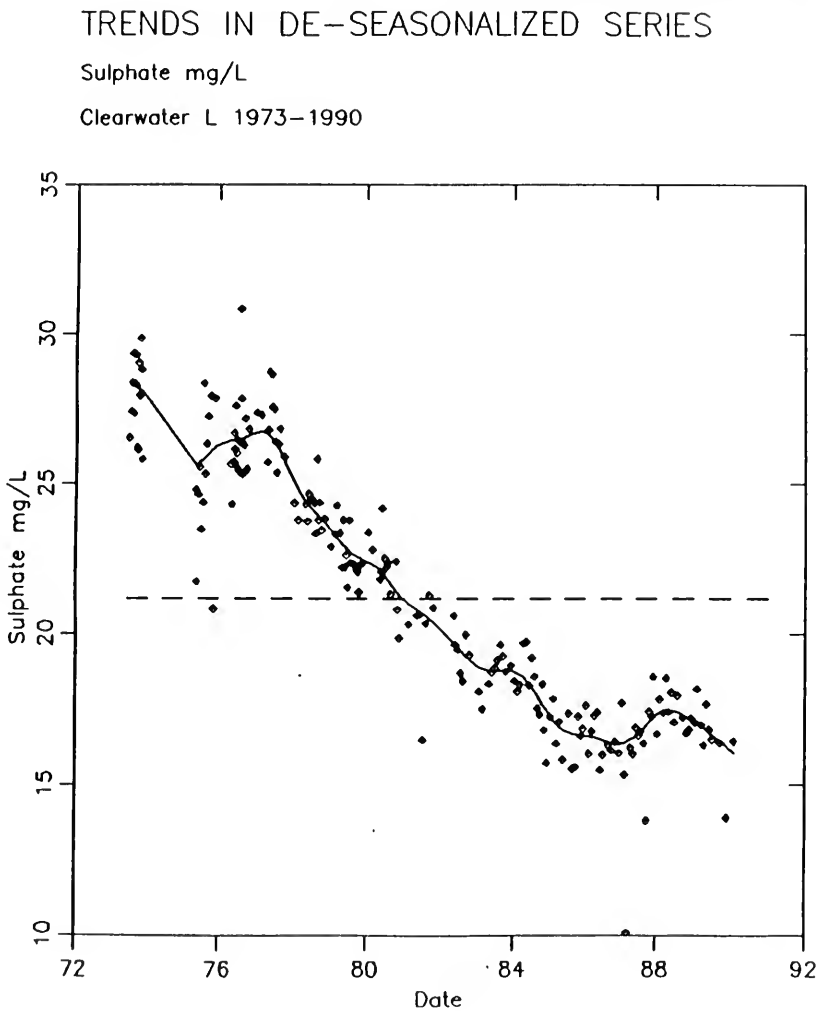


**Figure 16.1 Sulphate series in mg/L units.**




Run #1:
Run date: MAY 23 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 21.174

Maximum trend = 28.368
@ JUN 28 1973

Minimum trend = 16.036
@ FEB 5 1990

Trend range = 12.333

Observations = 202
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 5
Unit smoothing interval = 1 yr
Minimum smoothing on 14 observations

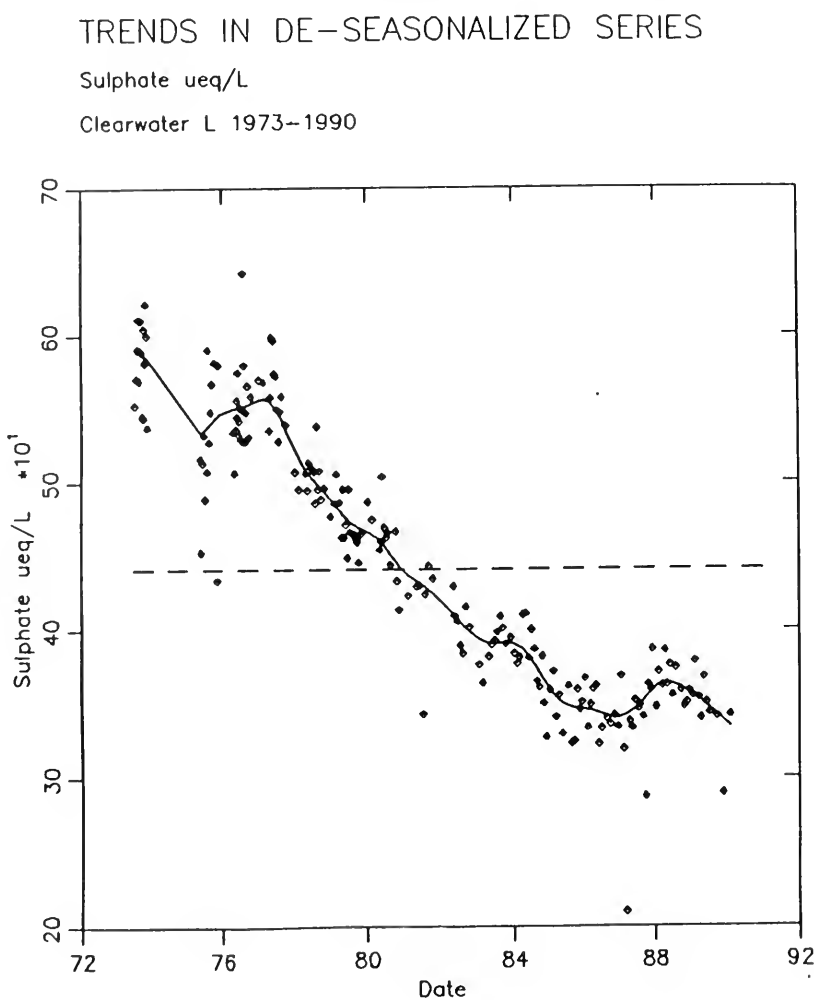
Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.962
p(RHO) = 0.000 dof = 14

% Variation due to trend = 92.8

Data file: SO4.DAT
Parameter: SO4
Site: Clearwater L

**Figure 16.2 Sulphate series in $\mu$eq/L units.**




Run #1:
Run date: MAY 23 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 440.859

Maximum trend = 590.638
@ JUN 28 1973

Minimum trend = 333.816
@ FEB 5 1990

Trend range = 256.821

Observations = 202
Series begins @ JUN 28 1973
Series ends @ FEB 5 1990

Filter = 5
Unit smoothing interval = 1 yr
Minimum smoothing on 14 observations

Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = -0.962
p(RHO) = 0.000 dof = 14

% Variation due to trend = 92.8

Data file: SO4.DAT
Parameter: SO4
Site: Clearwater L

**Figure 17.1 Single density plot.**

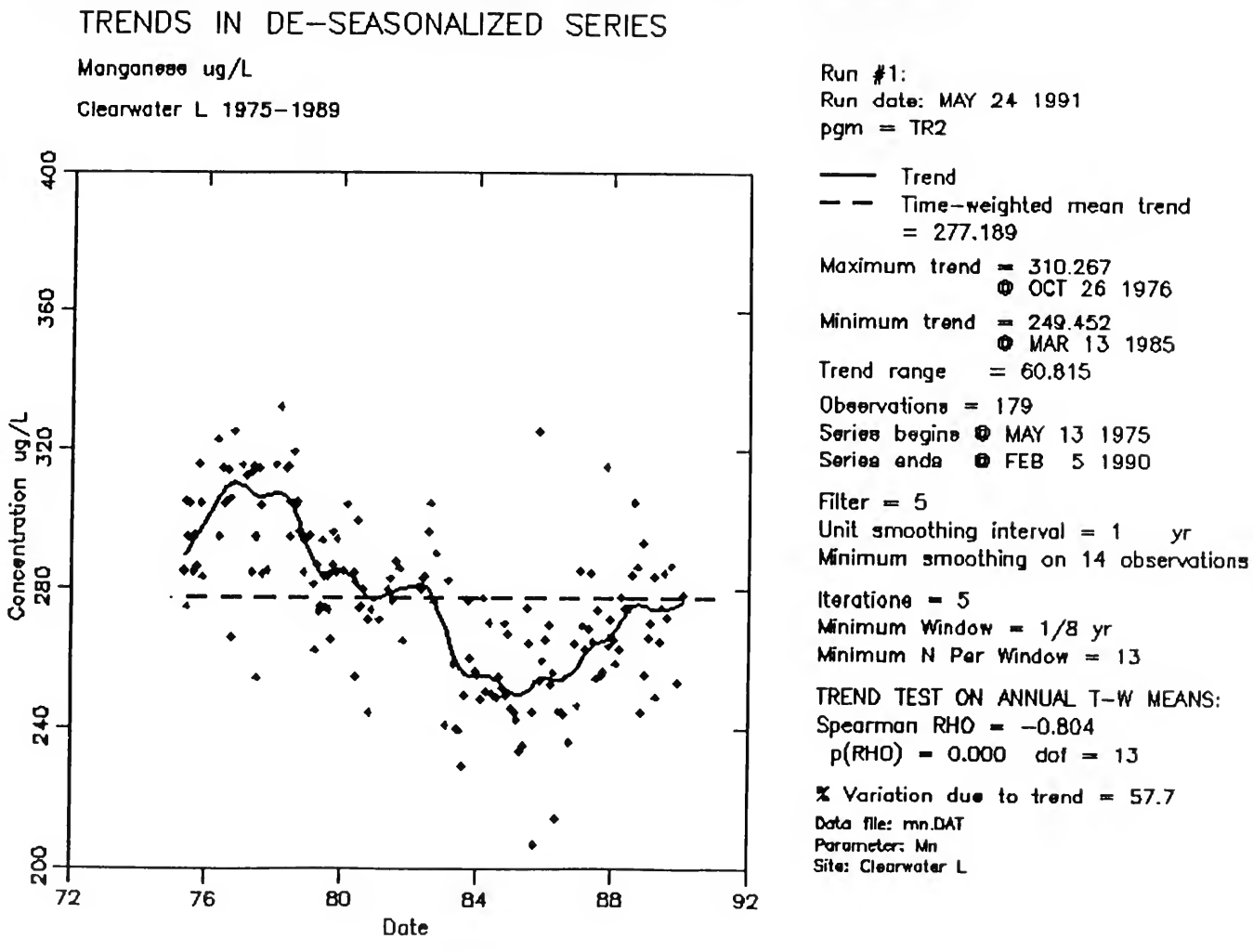
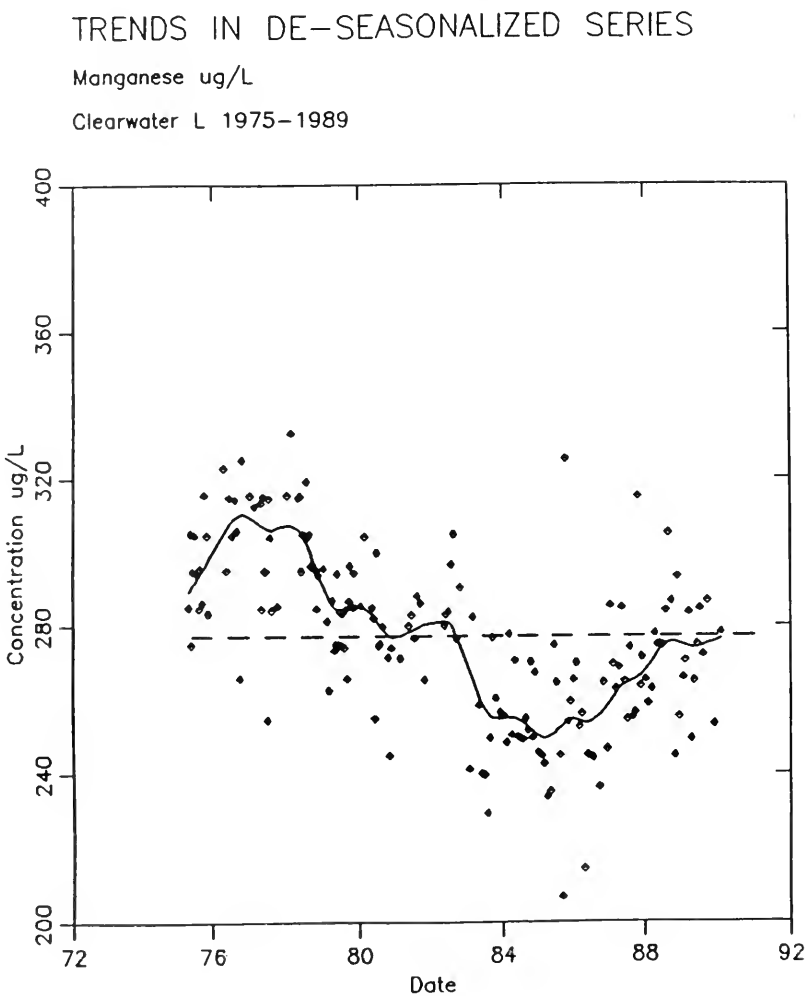


**Figure 17.2 Double density plot.**




Run #1:
Run date: MAY 24 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 277.189

Maximum trend = 310.267
@ OCT 26 1976

Minimum trend = 249.452
@ MAR 13 1985

Trend range = 60.815

Observations = 179
Series begins @ MAY 13 1975
Series ends @ FEB 5 1990

Filter = 5
Unit smoothing interval = 1 yr
Minimum smoothing on 14 observations

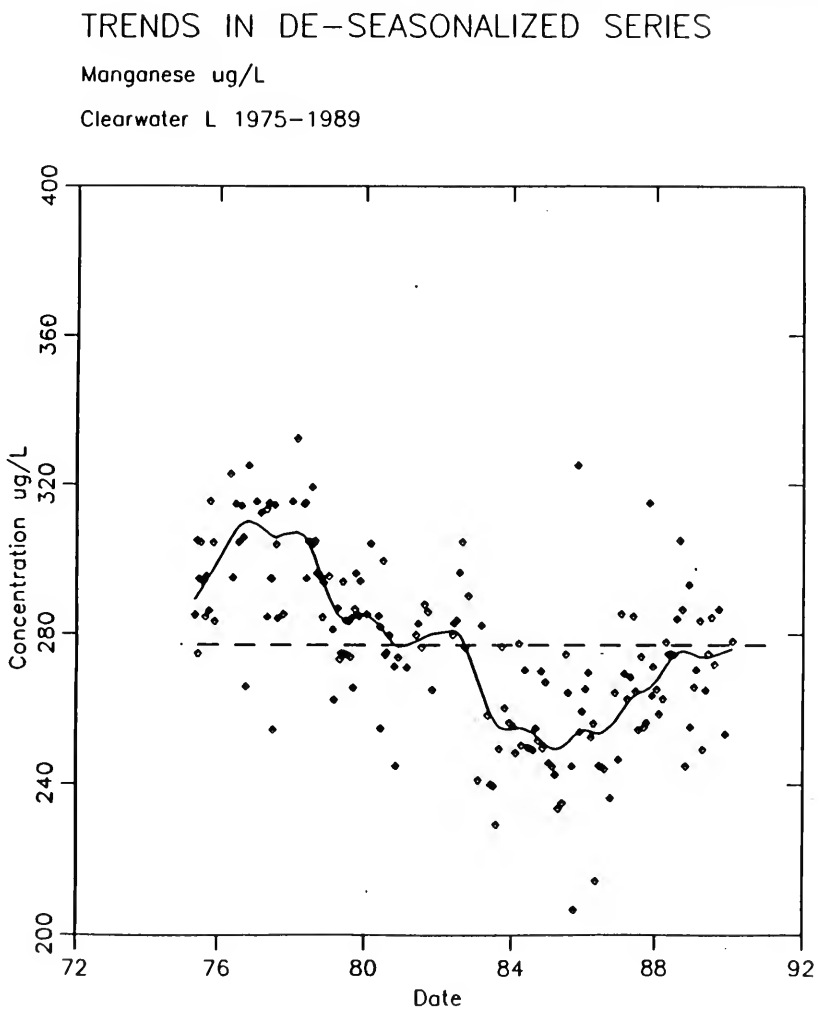
Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = −0.804
p(RHO) = 0.000 dof = 13

% Variation due to trend = 57.7

Data file: mn.DAT
Parameter: Mn
Site: Clearwater L

**Figure 17.3 Quad density plot.**




Run #1:
Run date: MAY 24 1991
pgm = TR2

——— Trend
— — Time-weighted mean trend
= 277.189

Maximum trend = 310.267
@ OCT 26 1976

Minimum trend = 249.452
@ MAR 13 1985

Trend range = 60.815

Observations = 179
Series begins @ MAY 13 1975
Series ends @ FEB 5 1990

Filter = 5
Unit smoothing interval = 1 yr
Minimum smoothing on 14 observations

Iterations = 5
Minimum Window = 1/8 yr
Minimum N Per Window = 13

TREND TEST ON ANNUAL T-W MEANS:
Spearman RHO = -0.804
p(RHO) = 0.000 dof = 13

% Variation due to trend = 57.7

Data file: mn.DAT
Parameter: Mn
Site: Clearwater L

**Figure 18.1 Normally scaled trend plot of $\log_{10}$ variates.**

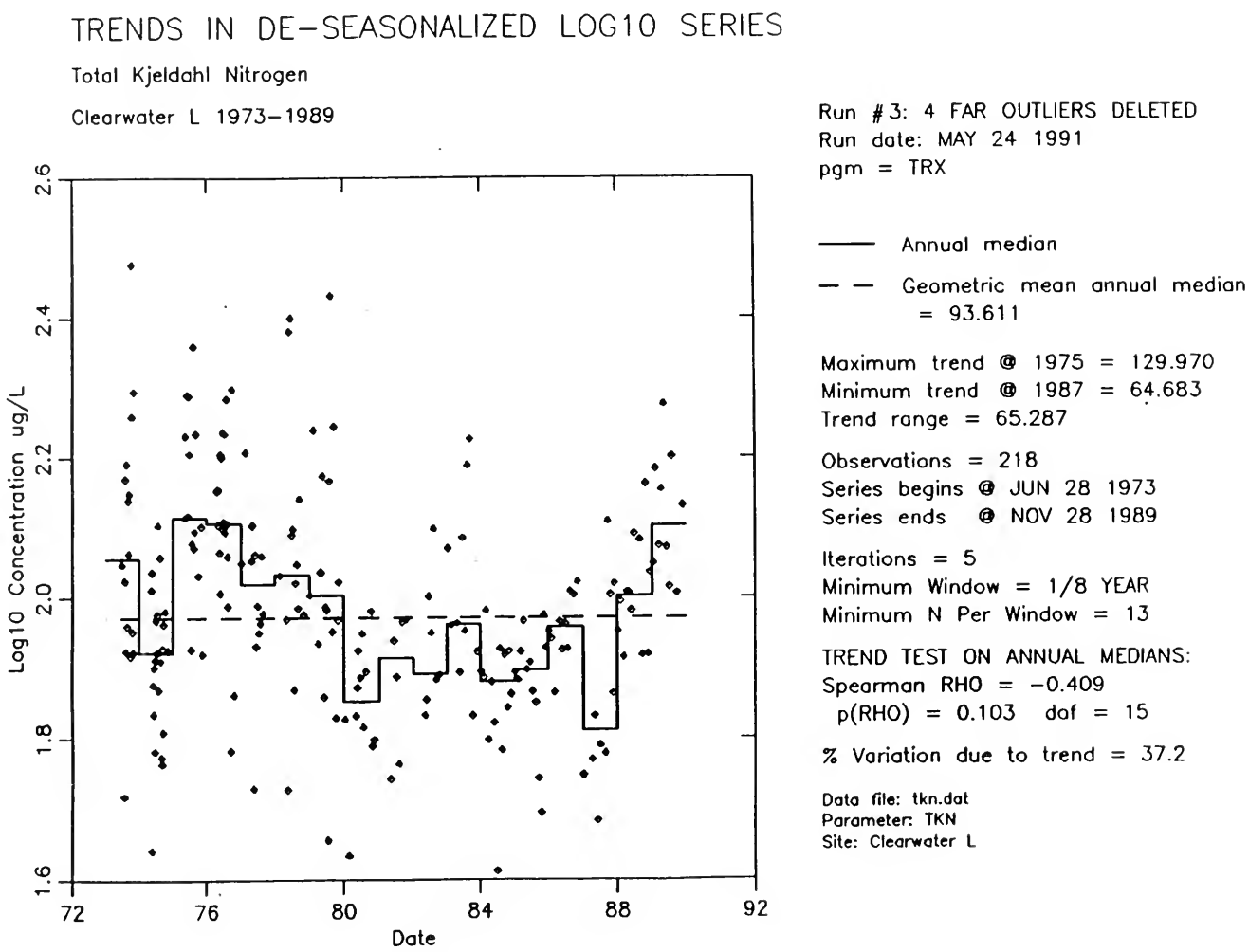


**Figure 18.2 Decadal style trend plot of $\log_{10}$ variates.**

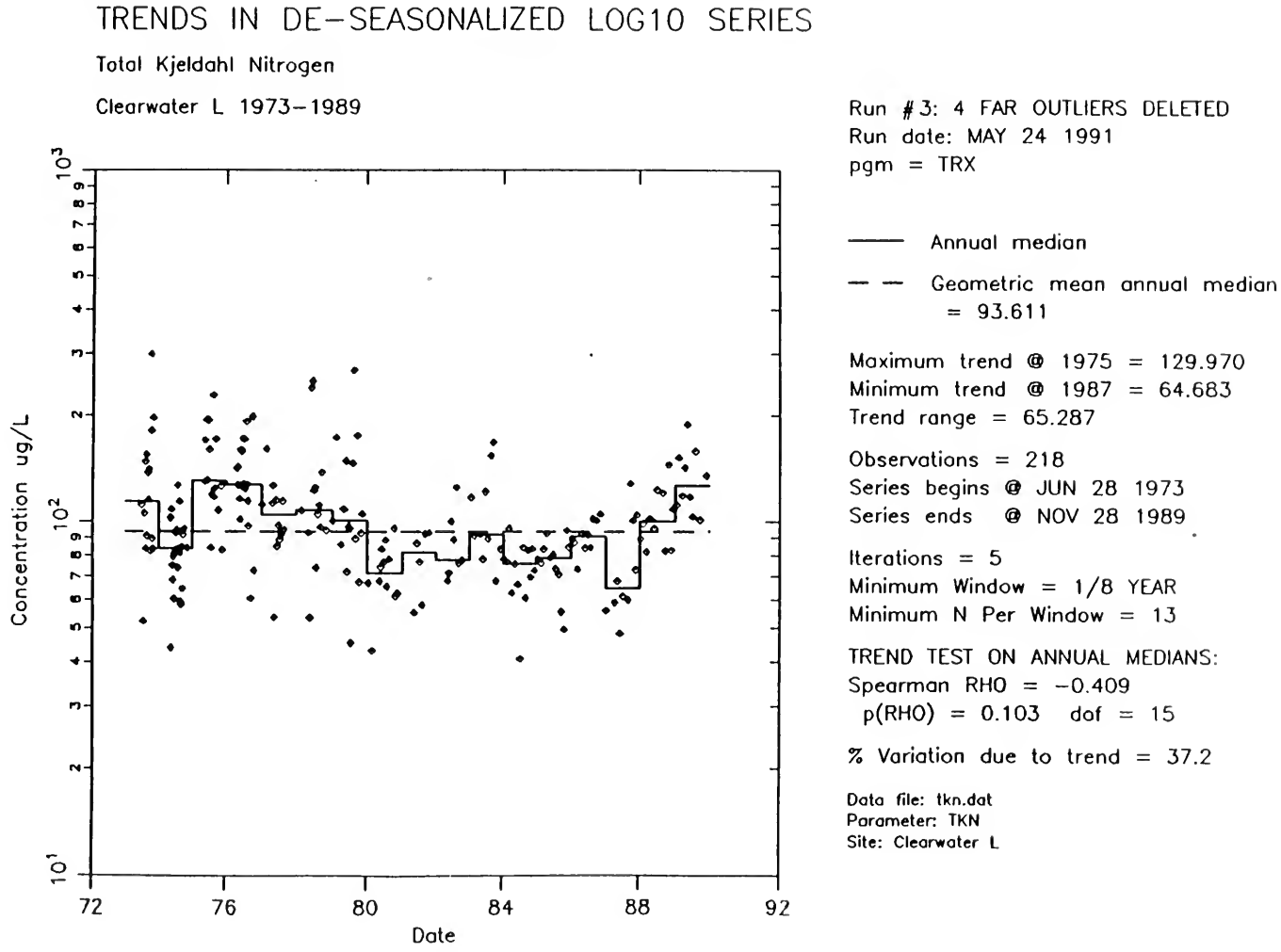


**Figure 18.3 Decadal style trend plot of $\log_{10}$ microbiological variates.**

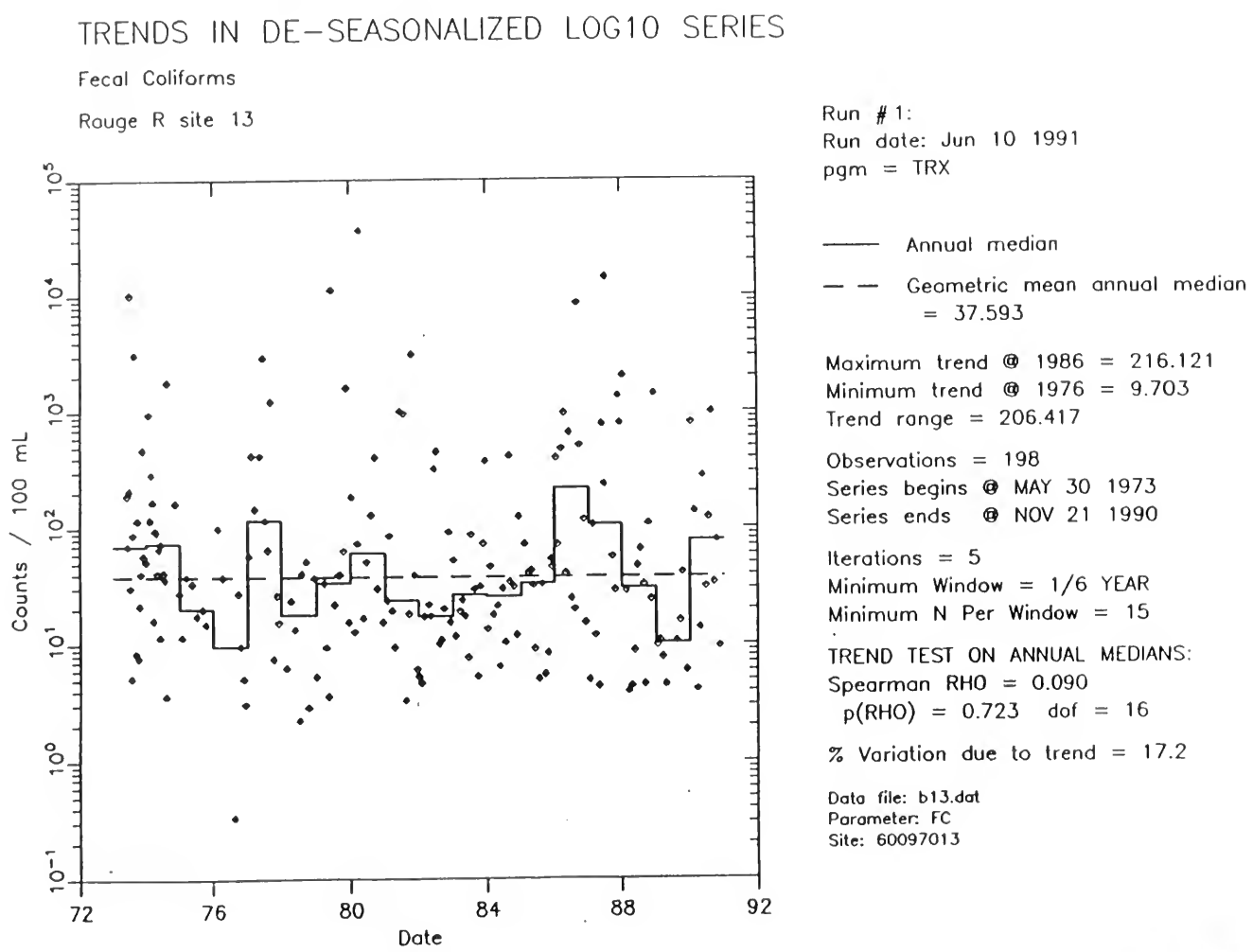


**Figure 19 'Short' plot without right side annotation.**

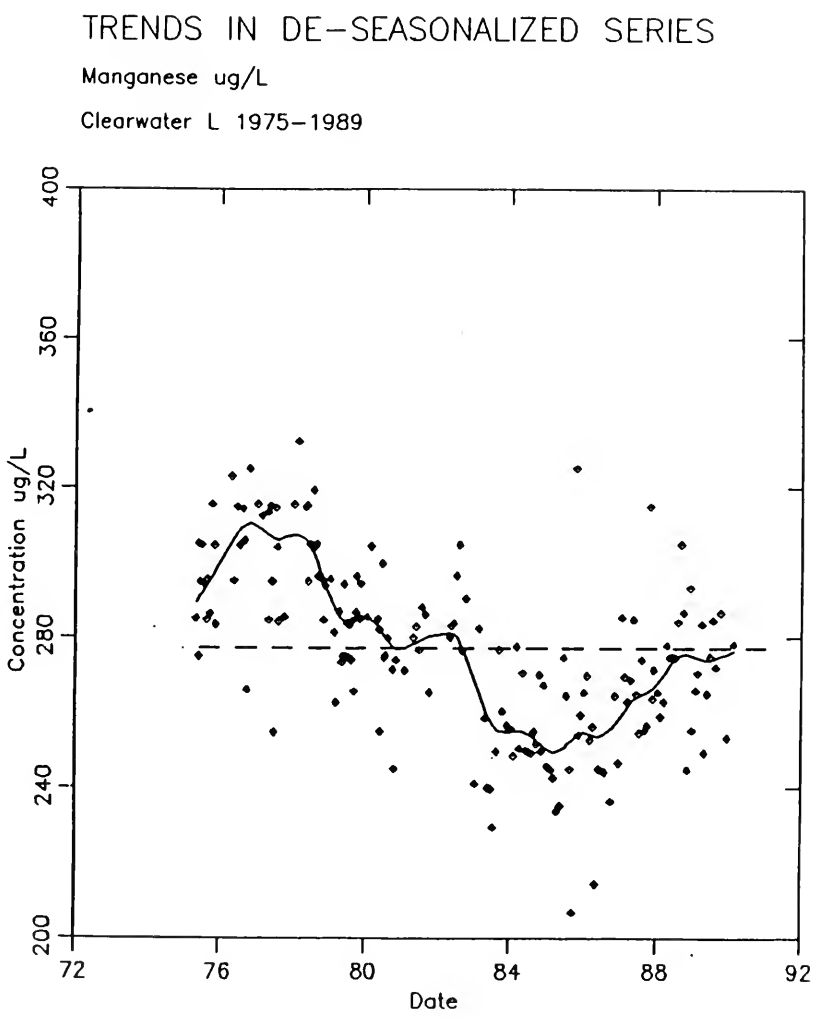


**Figure 20.1 Location of plot headers, date & $y$-axis title.**

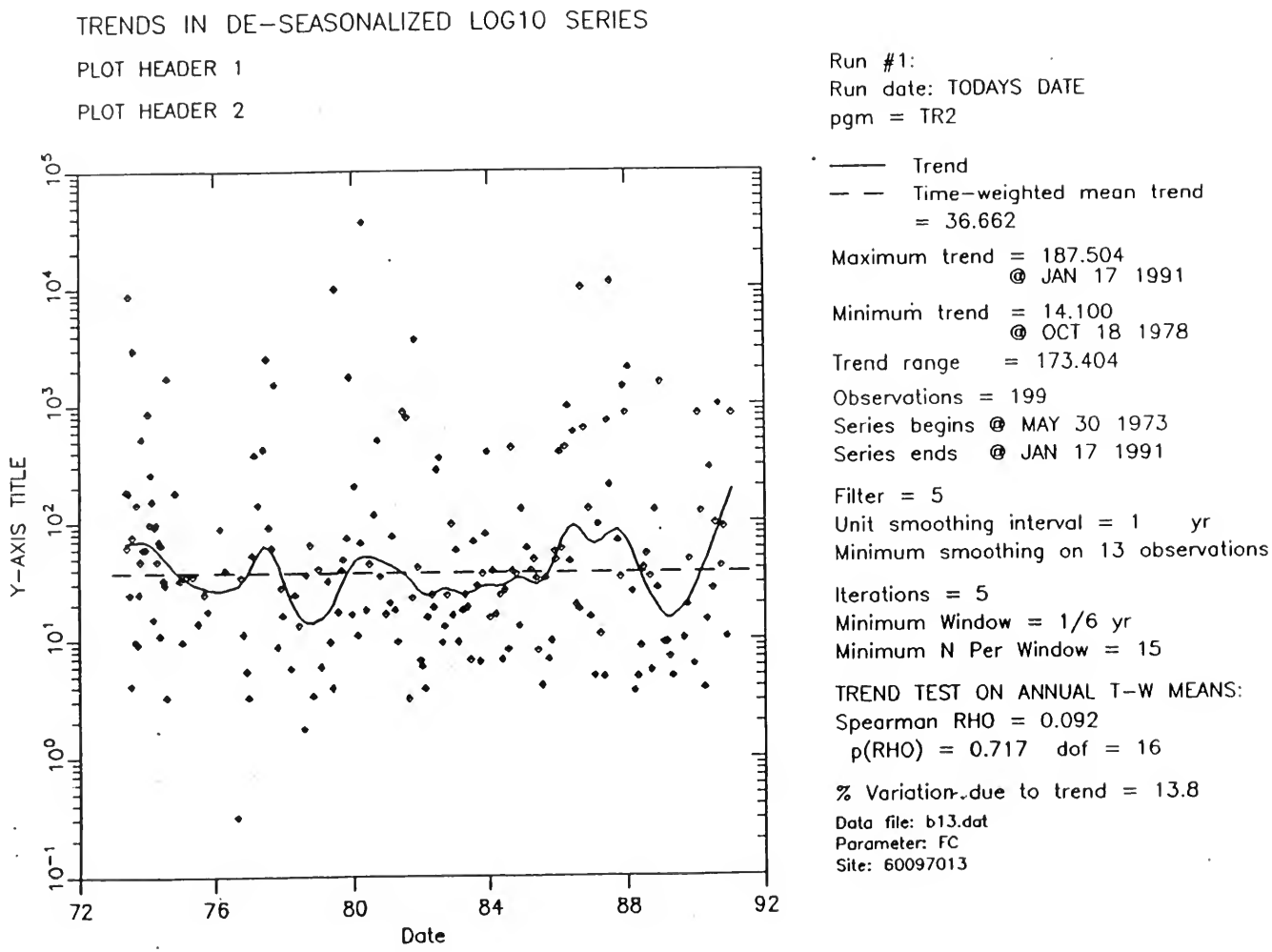
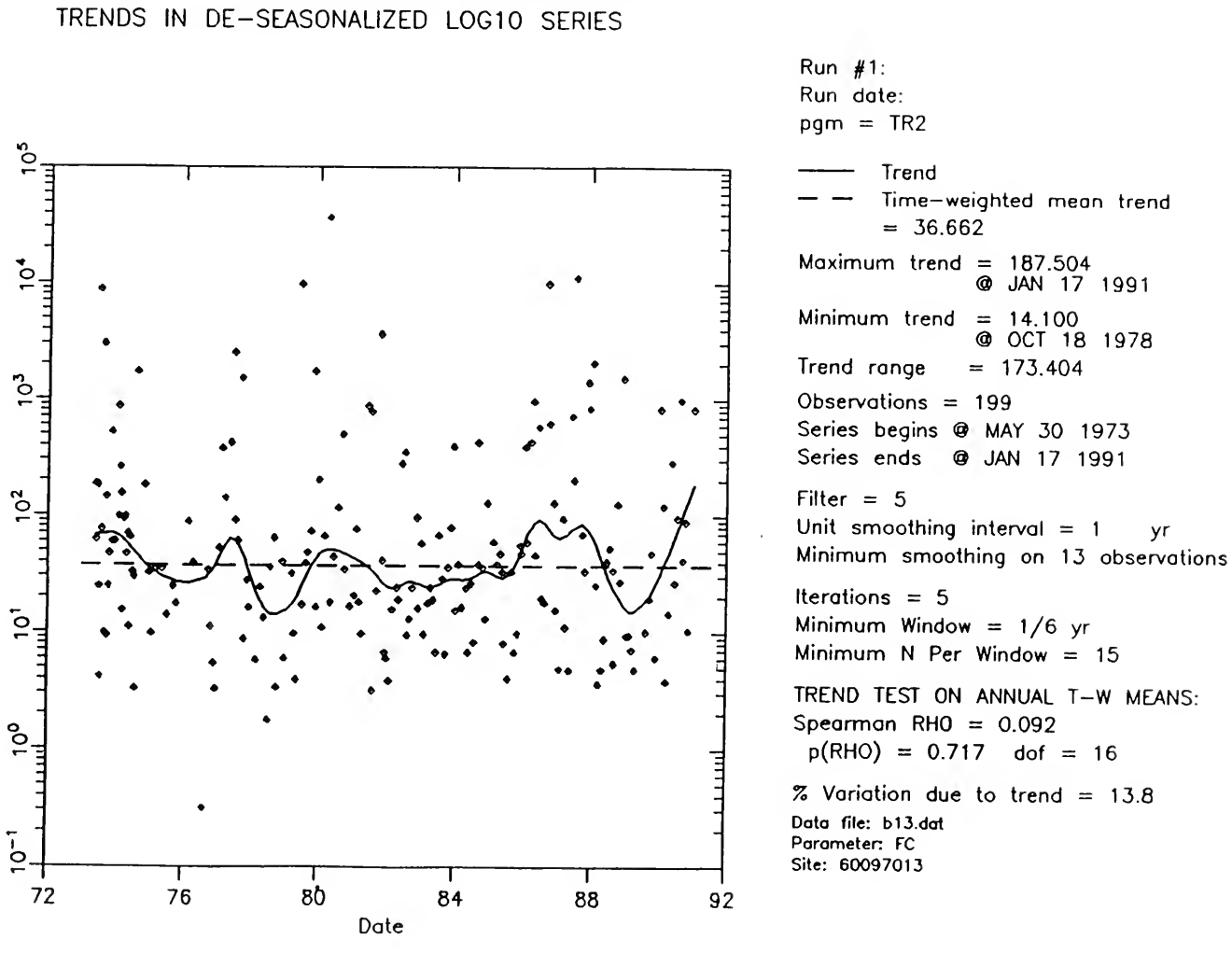


**Figure 20.2 Blank plot headers, date & $y$-axis title.**



# APPENDIX B

# TABLES

for

PC-Software, Users Guide and Documentation

for

Robust Graphical Time Series Analysis

of

Long Term Surface Water Quality Records

## APPENDIX: TABLE 1

**Table 1. Robust Smoothing Filters of Programs TR1 and TR2:**

**Filter 1:**

- basic LOWESS with 2 robustness iterations

**Filter 2:**

- LOWESP (modified LOWESS) with 2 robustness iterations

**Filter 3:**

3a. basic LOWESS with 2 robustness iterations

3b. basic LOWESS with 0 robustness iterations on output of 3a.

3c. basic LOWESS with 0 robustness iterations on output of 3b.

**Filter 4:**

4a. LOWESP with 2 robustness iterations

4b. LOWESP with 0 robustness iterations on output of 4a.

4c. LOWESP with 0 robustness iterations on output of 4b.

**Filter 5:**

5a. LOWESP with 2 robustness iterations.

5b. output of 5a is interpolated to an evenly spaced series.

5c. LOWESP with 2 robustness iterations on evenly spaced series 5b.

5d. output of 5c is interpolated to original irregular series.

5c. LOWESP with 0 robustness iterations on irregular series 5d.

**Filter 6:**

6a. LOWESP with 2 robustness iterations.

6b. LOWESP with 0 robustness iterations on output 6a as if series were evenly spaced.

6c. LOWESP with 0 robustness iterations on output 6b at original irregular spacing.

**Filter 7:**

7a. Tukey H3RSSH smooth as if series were evenly spaced

7b. Tukey H3RSSH smooth of 7a output as if series were evenly spaced

7c. LOWESP with 2 robustness iterations on output 7b at original irregular spacing.

7d. LOWESP with 0 robustness iterations on output 7c at original irregular spacing.

**Filter 8:**

8a. Tukey H3RSSH smooth as if series were evenly spaced

8b. Tukey H3RSSH smooth of 8a output as if series were evenly spaced

**Filter 9:**

- basic LOWESS with 0 robustness iterations

## APPENDIX: TABLE 2

Table 2. Seasonal Date Conversion Chart, (365 day year).

| Seasonal Date | Calendar Date | Time |
|---|---|---|
| .00 | JAN 1 | 00:00 |
| .01 | JAN 4 | 15:36 |
| .02 | JAN 8 | 07:12 |
| .03 | JAN 11 | 22:48 |
| .04 | JAN 15 | 14:24 |
| .05 | JAN 19 | 06:00 |
| .06 | JAN 22 | 21:36 |
| .07 | JAN 26 | 13:12 |
| .08 | JAN 30 | 04:48 |
| .09 | FEB 2 | 20:24 |
| .10 | FEB 6 | 12:00 |
| .11 | FEB 10 | 03:36 |
| .12 | FEB 13 | 19:12 |
| .13 | FEB 17 | 10:48 |
| .14 | FEB 21 | 02:24 |
| .15 | FEB 24 | 18:00 |
| .16 | FEB 28 | 09:36 |
| .17 | MAR 4 | 01:12 |
| .18 | MAR 7 | 16:48 |
| .19 | MAR 11 | 08:24 |
| .20 | MAR 15 | 00:00 |
| .21 | MAR 18 | 15:36 |
| .22 | MAR 22 | 07:12 |
| .23 | MAR 25 | 22:48 |
| .24 | MAR 29 | 14:24 |
| .25 | APR 2 | 06:00 |
| .26 | APR 5 | 21:36 |
| .27 | APR 9 | 13:12 |
| .28 | APR 13 | 04:48 |
| .29 | APR 16 | 20:24 |
| .30 | APR 20 | 12:00 |
| .31 | APR 24 | 03:36 |
| .32 | APR 27 | 19:12 |
| .33 | MAY 1 | 10:48 |
| .34 | MAY 5 | 02:24 |
| .35 | MAY 8 | 18:00 |
| .36 | MAY 12 | 09:36 |
| .37 | MAY 16 | 01:12 |
| .38 | MAY 19 | 16:48 |
| .39 | MAY 23 | 08:24 |

## APPENDIX: TABLE 2

Table 2. Seasonal Date Conversion Chart Cont'd, (365 day year).

| Seasonal Date | Calendar Date | Time |
|---|---|---|
| .40 | MAY 27 | 00:00 |
| .41 | MAY 30 | 15:36 |
| .42 | JUN 3 | 07:12 |
| .43 | JUN 6 | 22:48 |
| .44 | JUN 10 | 14:24 |
| .45 | JUN 14 | 06:00 |
| .46 | JUN 17 | 21:36 |
| .47 | JUN 21 | 13:12 |
| .48 | JUN 25 | 04:48 |
| .49 | JUN 28 | 20:24 |
| .50 | JUL 2 | 12:00 |
| .51 | JUL 6 | 03:36 |
| .52 | JUL 9 | 19:12 |
| .53 | JUL 13 | 10:48 |
| .54 | JUL 17 | 02:24 |
| .55 | JUL 20 | 18:00 |
| .56 | JUL 24 | 09:36 |
| .57 | JUL 28 | 01:12 |
| .58 | JUL 31 | 16:48 |
| .59 | AUG 4 | 08:24 |
| .60 | AUG 8 | 00:00 |
| .61 | AUG 11 | 15:36 |
| .62 | AUG 15 | 07:12 |
| .63 | AUG 18 | 22:48 |
| .64 | AUG 22 | 14:24 |
| .65 | AUG 26 | 06:00 |
| .66 | AUG 29 | 21:36 |
| .67 | SEP 2 | 13:12 |
| .68 | SEP 6 | 04:48 |
| .69 | SEP 9 | 20:24 |
| .70 | SEP 13 | 12:00 |
| .71 | SEP 17 | 03:36 |
| .72 | SEP 20 | 19:12 |
| .73 | SEP 24 | 10:48 |
| .74 | SEP 28 | 02:24 |
| .75 | OCT 1 | 18:00 |
| .76 | OCT 5 | 09:36 |
| .77 | OCT 9 | 01:12 |

## APPENDIX: TABLE 2

Table 2. Seasonal Date Conversion Chart Cont'd, (365 day year).

| Seasonal Date | Calendar Date | Time |
|---|---|---|
| .78 | OCT 12 | 16:48 |
| .79 | OCT 16 | 08:24 |
| .80 | OCT 20 | 00:00 |
| .81 | OCT 23 | 15:36 |
| .82 | OCT 27 | 07:12 |
| .83 | OCT 30 | 22:48 |
| .84 | NOV 3 | 14:24 |
| .85 | NOV 7 | 06:00 |
| .86 | NOV 10 | 21:36 |
| .87 | NOV 14 | 13:12 |
| .88 | NOV 18 | 04:48 |
| .89 | NOV 21 | 20:24 |
| .90 | NOV 25 | 12:00 |
| .91 | NOV 29 | 03:36 |
| .92 | DEC 2 | 19:12 |
| .93 | DEC 6 | 10:48 |
| .94 | DEC 10 | 02:24 |
| .95 | DEC 13 | 18:00 |
| .96 | DEC 17 | 09:36 |
| .97 | DEC 21 | 01:12 |
| .98 | DEC 24 | 16:48 |
| .99 | DEC 28 | 08:24 |
| 1.00 | DEC 31 | 24:00 |

## APPENDIX: TABLE 3

Table 3. Sample input data file.

```
----+----1----+----2----+----3
As    170026013
170026013 641117     0   320.
170026013 650308     0   130.
170026013 650323     0   200.
170026013 650405     0    50.
170026013 650412     0   200.
170026013 650531     0    80.
170026013 650617     0   300.
170026013 650621     0   300.
170026013 650708     0  1000.
170026013 650720     0  1500.
170026013 650803     0  1000.
170026013 650817     0  5000.
170026013 650818     0  5000.
170026013 650908     0  2500.
170026013 650916     0  4000.
170026013 650921     0  6000.
170026013 651021     0   250.
170026013 660830  1555  2500.
170026013 671002  1430     5.<
170026013 671101  1540   224.
170026013 680402  1125    40.
170026013 680522  1830    90.
170026013 680604     0   490.
170026013 680703     0    60.
170026013 680807     0   540.
170026013 680907     0   880.
170026013 681125  1600   290.
170026013 690217  1545    80.
170026013 690314  1530    60.
170026013 690422  1525    80.
170026013 690618  1415   142.
170026013 690715  1430   180.
170026013 690819  1440  1000.
170026013 690923  1430  1520.
170026013 691021  1410   470.
170026013 691111  1550   560.
```

Table 4. Sample input data file (matrix).

```
----+----1----+----2----+----3----+----4----+----5----+----6----+----7--

IONS  CLEARWATER L
 730628 1200      87.5000      62.5950    -999.0000    -999.0000    -999.0000
 730718 1200      91.0000      39.4948    -999.0000    -999.0000    -999.0000
 730802 1200      91.0000      49.7209    -999.0000    -999.0000    -999.0000
 730809 1200      86.5000      78.8024    -999.0000    -999.0000    -999.0000
 730814 1200      85.0000      49.7209    -999.0000    -999.0000    -999.0000
 730821 1200      86.0000      49.7209    -999.0000    -999.0000    -999.0000
 730828 1200      91.0000      49.7209    -999.0000    -999.0000    -999.0000
 730906 1200      99.0000      39.4948    -999.0000    -999.0000    -999.0000
 730912 1200      87.0000      47.9798    -999.0000    -999.0000    -999.0000
 730919 1200      92.0000      35.1998    -999.0000    -999.0000    -999.0000
 730925 1200      90.0000      49.7209    -999.0000    -999.0000    -999.0000
 731003 1200      90.0000      44.3139    -999.0000    -999.0000    -999.0000
 731010 1200      92.0000      49.7209    -999.0000    -999.0000    -999.0000
 731015 1200      91.0000      39.4948    -999.0000    -999.0000    -999.0000
 731023 1200      91.0000      49.7209    -999.0000    -999.0000    -999.0000
 731102 1200      89.3195      49.7209    -999.0000    -999.0000    -999.0000
 731106 1200      87.0000      78.8024    -999.0000    -999.0000    -999.0000
 740123 1200      93.0000      46.8597    -999.0000    -999.0000    -999.0000
 740219 1200      98.5000      46.5178    -999.0000    -999.0000    -999.0000
 740403 1200      97.0000      47.4875    -999.0000    -999.0000    -999.0000
 740515 1200      86.0000      41.8021    -999.0000    -999.0000    -999.0000
 740522 1200      87.5000      41.6212    -999.0000    -999.0000    -999.0000
 740528 1200      86.0000      41.5290    -999.0000    -999.0000    -999.0000
 740604 1200      82.0000      41.4964    -999.0000    -999.0000    -999.0000
 740612 1200      84.0000      41.6734    -999.0000    -999.0000    -999.0000
 740618 1200      84.0000      42.1103    -999.0000    -999.0000    -999.0000
 740625 1200      84.0000      43.1132    -999.0000    -999.0000    -999.0000
 740703 1200      84.0000      44.8428    -999.0000    -999.0000      67.4212
 740710 1200      85.5000      46.4174    -999.0000    -999.0000      63.0715
 740717 1200      85.5000      47.5864    -999.0000    -999.0000      60.8966
 740723 1200      84.0000      48.1177    -999.0000    -999.0000      67.4212
 740730 1200      81.6670      48.3363    -999.0000    -999.0000      65.2463
 740806 1200      85.0000      48.3447    -999.0000    -999.0000      58.7217
 740815 1200      84.6670      48.2714    -999.0000    -999.0000      65.2463
 740820 1200      87.0000      48.2371    -999.0000    -999.0000      63.0715
 740827 1200      86.0000      48.2061    -999.0000    -999.0000      58.7217
 740904 1200      87.5000      48.1900    -999.0000    -999.0000      52.1971
 740910 1200      87.5000      48.1830    -999.0000    -999.0000      55.4594
 740918 1200      85.0000      48.1626    -999.0000    -999.0000      73.9458
 740925 1200      88.5000      48.0769    -999.0000    -999.0000      73.9458
 741002 1200      88.0000      47.8517    -999.0000    -999.0000      71.7710
 741008 1200      87.5000      47.5405    -999.0000    -999.0000      67.4212
 741015 1200      86.5000      47.1130    -999.0000    -999.0000      65.2463
 741114 1200      85.5000      45.7974    -999.0000    -999.0000      58.7217
```

Table 5a. Sample .CNT file for program TRX

```
----+----1----+----2----+----3

D:\TRX\CRUD                     1:report file name
D:\moira\as011.dat              2:input data file name
As                              3:input variable identifier
(10X,I7,I5,f10.1)               4:input data format
 730000 900000                  5:bracket dates
1   -9.000                      6:zero option & missing value codes
1    1.000                      7:multiplier/divisor codes
  1  1  5                       8:log code, # runs, # iterations
 13  8                          9:seasonal smoothing parameters
 P S 1 L                       10:plot codes
Arsenic ug/L (<=DL)            11:plot header 1
Moira Hwy 62                   12:plot header 2
APR 22 1991                    13:current date
Concentration ug/L             14:y-axis title
SR                             15:long/short report code
1                              16:output file code
D:\TRX\CRUD1                   17:output trend file
D:\TRX\CRUD2                   18:output seasonal file
```

Table 5b. Sample .CNT file for program TR1

```
----+----1----+----2----+----3

D:\TR1\CRUD                     1:report file name
D:\clear\ion.dat                2:input data file name
SO4                             3:input variable identifier
(I7,I5,132X,f12.4)              4:input data format
 730000 910000                  5:bracket dates
1   -9.000                      6:zero option & missing value codes
M 20.82084                      7:multiplier/divisor codes
  0  3                          8:log code, # runs
 5 1  1.                        9:filter, sample interval type, USI
 P S 1 L                       10:plot codes
Sulphate ueq/L                 11:plot header 1
Clearwater L                   12:plot header 2
MAY 11 1991                    13:current date
Concentration ueq/L            14:y-axis title
SR                             15:long/short report code
1                              16:output file code
D:\TR1\CRUD1                   17:output trend file
```

Table 5c. Sample .CNT file for program TR2

```
----+----1----+----2----+----3

D:\TR2\CRUD                      1:report file name
D:\clear\cond.dat                2:input data file name
COND                             3:input variable identifier
(5X,I5,132X,f9.3)                4:input data format
 730000 910000                   5:bracket dates
1    -9.000                      6:zero option & missing value codes
1                                7:multiplier/divisor codes
  0  3  5                        8:log code, # runs
 5 1  1.                         9:filter, sample interval type, USI
 13  8                           10:seasonal smoothing parameters
 P Q 0 L                         11:plot codes
Conductance                      12:plot header 1
Clearwater L                     13:plot header 2
Jun 18 1991                      14:current date
Concentration uS/cm              15:y-axis title
SR                               16:long/short report code
1                                17:output file code
D:\TR2\CRUD1                     18:output trend file
D:\TR2\CRUD2                     19:output seasonal file
```

# APPENDIX C

# REPORT FILES

**for**

**PC-Software, Users Guide and Documentation**

**for**

**Robust Graphical Time Series Analysis**

**of**

**Long Term Surface Water Quality Records**

```
        PROGRAM TRX

 REPORT DATA FILE = D:\DILDO\TRX\CRNO3
 INPUT DATA FILE = D:\DILDO\DAT2\NO3.dat
 VARIABLE = NO3
(5X,I7,I5,F9.3)
 START YEAR = 720000   END YEAR = 900000
  ZERO OPTION CODE = 1
  NUMERIC MISSING VALUE CODE =    -9.00000000
  MULTIPLIER/DIVISOR CODE = 1
  MULTIPLIER/DIVISOR =       .00000000
  INPUT DATA ARE LEFT AS IS
  LOGG CODE  =   0
  RUNS =   1
  ITERATIONS =   5
  MINIMUM N / WINDOW =  13
  MINIMUM SEASONAL SMOOTHING WINDOW = 1/  8 yr
  PLOT CODES:
 PLOT INDICATOR = 0
 PLOT DENSITY 1 = D
 LOG PLOT TYPE  = 0
 PLOT TYPE CODE = L  (i.e. LONG/SHORT)
 PLOT HEADER 1 = Nitrate ug/L
 PLOT HEADER 2 = CLEARWATER L
 RUN DATE HEADER = JUN 16 1991
 Y-AXIS TITLE = Concentration ug/L
 REPORT FILE CODE = SR  (i.e. SHORT REPORT)
 OUTPUT FILE CODE = 0
 NO OUTPUT FILES REQUESTED
 FIRST LINE OF INPUT DATA FILE:
  PARAMETER = NO3       SITE = CLEARWATER L

  QUALITY DATA FOR PARM = NO3      READ & STORED
   212 RECORDS WERE READ

  THE FOLLOWING    3 INPUT DATA WERE IDENITIFIED AS ZERO'S:
     870901 1100       .000  ZERO
     880830 1030       .000  ZERO
     880926 1000       .000  ZERO

  SMALLEST NON-ZERO POSITIVE VALUE =     4.00000

  ZERO OPTION CODE = 1
  SETTING 0'S TO 1/2 SMALLEST POSITIVE NON-ZERO VALUE =     2.00000

*******************************************************************************
    BEGINNING RUN  1
```

TREND SUMMARY

| | YEAR | N | ITER = 1 | ITER = 5 | DIFF | YR-TO-YR | DIFF |
|---|---|---|---|---|---|---|---|
| 1 | 73 | 16 | 101.500 | 110.371 | -8.871 | 74-73 | -40.552 |
| 2 | 74 | 21 | 61.000 | 69.819 | -8.819 | 75-74 | -.764 |
| 3 | 75 | 10 | 66.000 | 69.055 | -3.055 | 76-75 | -12.524 |
| 4 | 76 | 13 | 46.000 | 56.531 | -10.531 | 77-76 | 3.996 |
| 5 | 77 | 10 | 63.500 | 60.528 | 2.972 | 78-77 | 26.366 |
| 6 | 78 | 14 | 85.000 | 86.894 | -1.894 | 79-78 | -15.616 |
| 7 | 79 | 19 | 70.000 | 71.278 | -1.278 | 80-79 | 5.982 |
| 8 | 80 | 12 | 72.500 | 77.260 | -4.760 | 81-80 | -6.458 |
| 9 | 81 | 7 | 60.000 | 70.802 | -10.802 | 82-81 | -5.471 |
| 10 | 82 | 7 | 55.000 | 65.331 | -10.331 | 83-82 | -28.902 |
| 11 | 83 | 10 | 38.000 | 36.429 | 1.571 | 84-83 | -13.586 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12 | 84 | 13 | 20.000 | 22.843 | -2.843 | 85-84 | 7.175 |
| 13 | 85 | 12 | 32.500 | 30.018 | 2.482 | 86-85 | 6.352 |
| 14 | 86 | 12 | 35.000 | 36.370 | -1.370 | 87-86 | -1.198 |
| 15 | 87 | 13 | 30.000 | 35.172 | -5.172 | 88-87 | -6.707 |
| 16 | 88 | 13 | 30.000 | 28.465 | 1.535 | 89-88 | .692 |
| 17 | 89 | 10 | 33.500 | 29.157 | 4.343 | | |

MEAN TREND = 52.912 56.254

MAXIMUM TREND = 110.371 @ 1973
MINIMUM TREND = 22.843 @ 1984
TREND RANGE = 87.528

SPEARMANS RHO TREND TEST ON FINAL ANNUAL TRENDS
FOR YEARS WITH WITH 5 OR MORE OBSERVATIONS
IF RHO < 0 THEN NET TREND IS DOWN
IF RHO > 0 THEN NET TREND IS UP
RHO = -.76471 APPROX p = .000350 dof = 15

**SSSSSSS SEASONAL SUMMARY SSSSSS**

MEAN ANNUAL SEASONAL MEDIAN = 56.548160

SCAN FOR SEASONAL EXTREMA

LOCAL MAX = 17.972840 @ .22 = MAR 22
LOCAL MIN = -11.976100 @ .62 = AUG 15

SEASONAL MAX = 17.972840 @ .22 = MAR 22
SEASONAL MIN = -11.976100 @ .62 = AUG 15
SEASONAL AMPLITUDE = 29.948940

ITERATIONS = 5 COMPLETED

**SSS VARIATION IN DE-SEASONALIZED RESIDUALS SSS**

VARIATION DUE TO TREND = 70.15365 %
VARIATION DUE TO NOISE = 29.84635 %
APPROXIMATE RSQ = 70.15365 %

| | | | |
|---|---|---|---|
| MEAN X = | 59.029960 | MEAN E = | 1.056031 |
| STD DEV X = | 30.731770 | STD DEV E = | 16.824720 |

ROBUST SUMMARY STATISTICS:

| | | | |
|---|---|---|---|
| MEDIAN X = | 56.793370 | MEDIAN E = | .140057 |
| MEDIAN \|X\| = | 56.793370 | MEDIAN \|E\| = | 8.238347 |
| MAD X = | 23.328770 | MAD E = | 8.174455 |
| .5 X IQR = | 23.284600 | .5 E IQR = | 8.595749 |
| R STD DEV = | 34.587290 | R STD DEV = | 12.119460 |

N.B. X = DE-SESEASONALIZED RESIDUAL
N.B. T = TREND COMPONENT
N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL

$$$ VARIATION IN DE-TRENDED RESIDUALS $$$

| | |
|---|---|
| VARIATION DUE TO SEASONAL | = 24.71026 % |
| VARIATION DUE TO NOISE | = 75.28974 % |
| APPROXIMATE RSQ | = 24.71026 % |

| | | | | | |
|---|---|---|---|---|---|
| MEAN X | = | 53.931260 | MEAN E | = | .762146 |
| STD DEV X | = | 19.232020 | STD DEV E | = | 16.824720 |

ROBUST SUMMARY STATISTICS:

| | | | | | |
|---|---|---|---|---|---|
| MEDIAN X | = | 53.702680 | MEDIAN E | = | -.153827 |
| MEDIAN \|X\| | = | 53.702680 | MEDIAN \|E\| | = | 8.319060 |
| MAD X | = | 10.102360 | MAD E | = | 8.174457 |
| .5 X IQR | = | 10.130100 | .5 E IQR | = | 8.595750 |
| R STD DEV | = | 14.977780 | R STD DEV | = | 12.119470 |

N.B. X = DE-TRENDED RESIDUAL
N.B. S = SEASONAL COMPONENT
N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL

$$$ TOTAL VARIATION IN SERIES: $$$

| | |
|---|---|
| VARIATION DUE TO TREND | = 58.21906 % |
| VARIATION DUE TO SEASON | = -5.71735 % |
| VARIATION DUE TO S+T | = 68.45300 % |
| VARIATION DUE TO NOISE | = 31.54700 % |
| APPROXIMATE RSQ | = 68.45300 % |

| | | | | | |
|---|---|---|---|---|---|
| MEAN X | = | 55.650940 | MEAN E | = | .909089 |
| STD DEV X | = | 29.989430 | STD DEV E | = | 16.824720 |

ROBUST SUMMARY STATISTICS

| | | | | | |
|---|---|---|---|---|---|
| MEDIAN X | = | 51.000000 | MEDIAN E | = | -.006885 |
| MEDIAN \|X\| | = | 51.000000 | MEDIAN \|E\| | = | 8.181341 |
| MAD X | = | 20.500000 | MAD E | = | 8.174456 |
| .5 X IQR | = | 20.281250 | .5 E IQR | = | 8.595751 |
| R STD DEV | = | 30.393350 | R STD DEV | = | 12.119470 |

N.B. X = RAW DATA
N.B. T = TREND COMPONENT
N.B. S = SEASONAL COMPONENT
N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL
N.B. D = CHRONOLOGICAL SAMPLE DATE

SUMMARY STATISTICS FOR DE-SEASONALIZED, DE-TRENDED RESIDUALS *****

| | |
|---|---|
| OBS | 212 |
| A-MEAN | .9091 |
| STD DEV | 16.8247 |
| MAX | 87.5349 |
| Q90 | 18.3060 |
| Q75 | 9.6436 |

| | |
|---|---|
| Q50 | -.0069 |
| Q25 | -7.5479 |
| Q10 | -16.4565 |
| MIN | -82.4167 |

| RANGE STATS | |
|---|---|
| MAX-MIN | 169.9516 |
| Q90-Q10 | 34.7625 |
| Q75-Q25 | 17.1915 |

TUKEY FENCES (STANDARDIZED) FOR OUTLIER SCAN

| | | |
|---|---|---|
| UPPER OUTER FENCE = | 61.2181 | 7.4826 |
| UPPER INNER FENCE = | 35.4308 | 4.3307 |
| LOWER INNER FENCE = | -33.3352 | -4.0745 |
| LOWER OUTER FENCE = | -59.1225 | -7.2265 |

!!! OUTLIERS IN DE-TRENDED & DE-SEASONALIZED RESIDUALS !!!

| | DATE | TIME | RAW DATUM | RESIDUAL | STANDARDIZED | |
|---|---|---|---|---|---|---|
| RESIDUAL | 730809 | 1200 | 162.000 | 63.396 | 7.749 | IS FAR OUT |
| RESIDUAL | 741114 | 1200 | 151.000 | 87.535 | 10.699 | IS FAR OUT |
| RESIDUAL | 751014 | 1200 | 91.000 | 30.704 | 3.753 | IS ALMOST OUT |
| RESIDUAL | 760511 | 1200 | 126.000 | 64.465 | 7.880 | IS FAR OUT |
| RESIDUAL | 760607 | 1200 | 19.000 | -35.582 | -4.349 | IS OUT |
| RESIDUAL | 780228 | 1200 | 20.000 | -82.417 | -10.074 | IS FAR OUT |
| RESIDUAL | 790219 | 1200 | 121.000 | 37.267 | 4.555 | IS OUT |
| RESIDUAL | 790309 | 1200 | 128.000 | 39.319 | 4.806 | IS OUT |
| RESIDUAL | 800730 | 1055 | 14.000 | -51.922 | -6.346 | IS ALMOST FAR OUT |
| RESIDUAL | 820914 | 830 | 24.000 | -30.819 | -3.767 | IS ALMOST OUT |
| RESIDUAL | 821018 | 1130 | 20.000 | -36.830 | -4.502 | IS OUT |
| RESIDUAL | 870512 | 1200 | 76.000 | 35.952 | 4.394 | IS OUT |

THERE ARE:
- 2 ALMOST OUT DATA
- 5 OUT DATA
- 1 ALMOST FAR OUT DATA
- 4 FAR OUT DATA
- 12 TOTAL OUTLIERS IDENTIFIED

N.B. OUTLIERS ARE STANDARDIZED BY DIVISION BY THE MEDIAN ABSOLUTE RESIDUAL = S

FOR NORMAL DISTRIBUTIONS:
OUTER FENCES ARE ABOUT 7S OR 4.7 X STANDARD DEVIATION
INNER FENCES ARE ABOUT 4S OR 2.7 X STANDARD DEVIATION
ALSO 6S OR 4 X STANDARD DEVIATION (ALMOST FAR OUT)
& 3.75S OR 2.5 X STANDARD DEVIATION (ALMOST OUT)
ARE STANDARDIZED VALUES USE BY OTHER AUTHORS

N.B. TUKEY FENCES ARE DEFINED USING 1/2 THE INTERQUARTILE RANGE
OF RESIDUALS WHICH IS NOT EXACTLY IDENTICAL TO S
EXCEPT FOR NORMAL DISTRIBUTIONS BUT IS ALSO A COMMON
ROBUST MEASURE OF DISPERSION

```
        PROGRAM TRX

 REPORT DATA FILE = D:\DILDO\TRX\CRNO3L
 INPUT DATA FILE = D:\DILDO\DAT2\NO3.dat
 VARIABLE = NO3
(5X,i7,i5,f9.3)
 START YEAR = 720000   END YEAR = 900000
  ZERO OPTION CODE = 1
  NUMERIC MISSING VALUE CODE =    -9.00000000
  MULTIPLIER/DIVISOR CODE = 1
  MULTIPLIER/DIVISOR =       .00000000
  INPUT DATA ARE LEFT AS IS
  LOGG CODE  =   1
  RUNS =   1
  ITERATIONS =   5
  MINIMUM N / WINDOW =  13
  MINIMUM SEASONAL SMOOTHING WINDOW = 1/  8 yr
  PLOT CODES:
 PLOT INDICATOR = 0
 PLOT DENSITY 1 = 0
 LOG PLOT TYPE  = 0
 PLOT TYPE CODE = L  (i.e. LONG/SHORT)
 PLOT HEADER 1 = Nitrate ug/L
 PLOT HEADER 2 = CLEARWATER L
 RUN DATE HEADER = JUN 13 1991
 Y-AXIS TITLE = Log10 Concentration ug/L
 REPORT FILE CODE = SR  (i.e. SHORT REPORT)
 OUTPUT FILE CODE = 0
 NO OUTPUT FILES REQUESTED
 FIRST LINE OF INPUT DATA FILE:
  PARAMETER = NO3       SITE = CLEARWATER L

  QUALITY DATA FOR PARM = NO3      READ & STORED
   212 RECORDS WERE READ

  THE FOLLOWING    3 INPUT DATA WERE IDENITIFIED AS ZERO'S:
     870901 1100       .000  ZERO
     880830 1030       .000  ZERO
     880926 1000       .000  ZERO

  SMALLEST NON-ZERO POSITIVE VALUE =      4.00000

  ZERO OPTION CODE = 1
  SETTING 0'S TO 1/2 SMALLEST POSITIVE NON-ZERO VALUE =      2.00000


  DOING LOG10 TRANSFORM OF DATA SERIES

*******************************************************************************
    BEGINNING RUN  1


                TREND SUMMARY ON LOG10 SERIES
        YEAR    N    ITER = 1   ITER = 5      DIFF    YR-TO-YR      DIFF
   1    73     16       2.006      2.084     -.077     74-73       -.222
   2    74     21       1.785      1.862     -.077     75-74       -.009
   3    75     10       1.818      1.853     -.035     76-75       -.097
   4    76     13       1.663      1.756     -.093     77-76        .032
   5    77     10       1.803      1.788      .014     78-77        .186
   6    78     14       1.929      1.975     -.045     79-78       -.120
   7    79     19       1.845      1.855     -.010     80-79        .050
   8    80     12       1.860      1.905     -.045     81-80       -.039
```

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 9 | 81 | 7 | 1.778 | 1.866 | -.088 | 82-81 | -.034 |
| 10 | 82 | 7 | 1.740 | 1.832 | -.092 | 83-82 | -.263 |
| 11 | 83 | 10 | 1.578 | 1.569 | .009 | 84-83 | -.293 |
| 12 | 84 | 13 | 1.301 | 1.277 | .024 | 85-84 | .185 |
| 13 | 85 | 12 | 1.511 | 1.462 | .049 | 86-85 | .094 |
| 14 | 86 | 12 | 1.544 | 1.556 | -.012 | 87-86 | -.025 |
| 15 | 87 | 13 | 1.477 | 1.531 | -.054 | 88-87 | -.082 |
| 16 | 88 | 13 | 1.477 | 1.449 | .028 | 89-88 | -.015 |
| 17 | 89 | 10 | 1.523 | 1.435 | .088 | | |

MEAN TREND = 1.685 1.709

MAXIMUM TREND = 2.084 @ 1973
MINIMUM TREND = 1.277 @ 1984
TREND RANGE = .807

TREND SUMMARY ON NATURAL SERIES

| | YEAR | N | ITER = 1 | ITER = 5 | DIFF | YR-TO-YR | DIFF |
|---|---|---|---|---|---|---|---|
| 1 | 73 | 16 | 101.499 | 121.323 | -19.825 | 74-73 | -48.491 |
| 2 | 74 | 21 | 61.000 | 72.832 | -11.832 | 75-74 | -1.545 |
| 3 | 75 | 10 | 65.810 | 71.287 | -5.477 | 76-75 | -14.303 |
| 4 | 76 | 13 | 46.000 | 56.984 | -10.984 | 77-76 | 4.420 |
| 5 | 77 | 10 | 63.482 | 61.403 | 2.079 | 78-77 | 32.896 |
| 6 | 78 | 14 | 85.000 | 94.299 | -9.299 | 79-78 | -22.690 |
| 7 | 79 | 19 | 70.000 | 71.610 | -1.610 | 80-79 | 8.742 |
| 8 | 80 | 12 | 72.457 | 80.352 | -7.895 | 81-80 | -6.826 |
| 9 | 81 | 7 | 60.000 | 73.526 | -13.526 | 82-81 | -5.603 |
| 10 | 82 | 7 | 55.000 | 67.923 | -12.923 | 83-82 | -30.815 |
| 11 | 83 | 10 | 37.881 | 37.108 | .773 | 84-83 | -18.187 |
| 12 | 84 | 13 | 20.000 | 18.921 | 1.079 | 85-84 | 10.029 |
| 13 | 85 | 12 | 32.404 | 28.950 | 3.453 | 86-85 | 7.016 |
| 14 | 86 | 12 | 35.000 | 35.966 | -.966 | 87-86 | -2.006 |
| 15 | 87 | 13 | 30.000 | 33.960 | -3.960 | 88-87 | -5.814 |
| 16 | 88 | 13 | 30.000 | 28.146 | 1.854 | 89-88 | -.942 |
| 17 | 89 | 10 | 33.317 | 27.204 | 6.112 | | |

G-MEAN TREND = 48.382 51.183

MAXIMUM TREND = 121.323 @ 1973
MINIMUM TREND = 18.921 @ 1984
TREND RANGE = 102.403

SPEARMANS RHO TREND TEST ON FINAL ANNUAL TRENDS
FOR YEARS WITH WITH 5 OR MORE OBSERVATIONS
IF RHO < 0 THEN NET TREND IS DOWN
IF RHO > 0 THEN NET TREND IS UP
RHO = -.77451 APPROX p = .000262 dof = 15

$$$$$$$ SEASONAL SUMMARY $$$$$$

MEAN ANNUAL SEASONAL MEDIAN = 1.710786 LOG UNITS
G-MEAN ANNUAL SEASONAL MEDIAN = 51.379040 NATURAL UNITS

SCAN FOR SEASONAL EXTREMA

LOCAL MAX = .176214 @ .25 = APR 2
LOCAL MIN = -.106097 @ .62 = AUG 15

SEASONAL MAX = .176214 @ .25 = APR 2

```
SEASONAL MIN =    -.106097 @  .62  =  AUG 15
SEASONAL AMPLITUDE =     .282311


 IN NATURAL UNITS
 AT G-MEAN ANNUAL SEASONAL MEDIAN =     51.379

SEASONAL MAX =   77.090340 @  .25  =  APR  2
SEASONAL MIN =   40.242840 @  .62  =  AUG 15
SEASONAL AMPLITUDE =   36.847500


 AT MAXIMUM TREND =    121.323  IN 1973

SEASONAL MAX =  182.036500 @  .25  =  APR  2
SEASONAL MIN =   95.027050 @  .62  =  AUG 15
SEASONAL AMPLITUDE =   87.009480


 AT MINIMUM TREND =     18.921   IN 1984

SEASONAL MAX =   28.389340 @  .25  =  APR  2
SEASONAL MIN =   14.819860 @  .62  =  AUG 15
SEASONAL AMPLITUDE =   13.569480

 ITERATIONS =  5  COMPLETED


$$$  VARIATION IN DE-SEASONALIZED RESIDUALS  $$$

   VARIATION DUE TO TREND    =  55.40270 %
   VARIATION DUE TO NOISE    =  44.59730 %
   APPROXIMATE RSQ           =  55.40270 %


   MEAN X      =     1.689344     MEAN E      =     -.031989
   STD DEV X   =      .324584     STD DEV E   =      .214785


   ROBUST SUMMARY STATISTICS:

   MEDIAN X    =     1.750334     MEDIAN E    =      .000181
   MEDIAN |X|  =     1.750334     MEDIAN |E|  =      .075157
   MAD X       =      .181839     MAD E       =      .075277
   .5 X IQR    =      .196041     .5 E IQR    =      .075290
   R STD DEV   =      .269595     R STD DEV   =      .111606

   N.B. X = DE-SESEASONALIZED RESIDUAL
   N.B. T = TREND COMPONENT
   N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL



$$$  VARIATION IN DE-TRENDED RESIDUALS  $$$

   VARIATION DUE TO SEASONAL    =  23.71050 %
   VARIATION DUE TO NOISE       =  76.28950 %
   APPROXIMATE RSQ              =  23.71050 %


   MEAN X      =     1.645320     MEAN E      =     -.033648
   STD DEV X   =      .240116     STD DEV E   =      .214785
```

ROBUST SUMMARY STATISTICS:

| | | | | | |
|---|---|---|---|---|---|
| MEDIAN X | = | 1.664036 | MEDIAN E | = | -.001478 |
| MEDIAN \|X\| | = | 1.664036 | MEDIAN \|E\| | = | .075868 |
| MAD X | = | .092489 | MAD E | = | .075277 |
| .5 X IQR | = | .092368 | .5 E IQR | = | .075291 |
| R STD DEV | = | .137125 | R STD DEV | = | .111606 |

N.B. X = DE-TRENDED RESIDUAL
N.B. S = SEASONAL COMPONENT
N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL

$$$ TOTAL VARIATION IN SERIES: $$$

| | | |
|---|---|---|
| VARIATION DUE TO TREND | = | 43.20377 % |
| VARIATION DUE TO SEASON | = | 2.86671 % |
| VARIATION DUE TO S+T | = | 56.64535 % |
| VARIATION DUE TO NOISE | = | 43.35465 % |
| APPROXIMATE RSQ | = | 56.64535 % |

| | | | | | |
|---|---|---|---|---|---|
| MEAN X | = | 1.657526 | MEAN E | = | -.032818 |
| STD DEV X | = | .325794 | STD DEV E | = | .214785 |

ROBUST SUMMARY STATISTICS

| | | | | | |
|---|---|---|---|---|---|
| MEDIAN X | = | 1.707570 | MEDIAN E | = | -.000649 |
| MEDIAN \|X\| | = | 1.707570 | MEDIAN \|E\| | = | .075251 |
| MAD X | = | .167491 | MAD E | = | .075277 |
| .5 X IQR | = | .169037 | .5 E IQR | = | .075291 |
| R STD DEV | = | .248323 | R STD DEV | = | .111606 |

N.B. X = RAW DATA
N.B. T = TREND COMPONENT
N.B. S = SEASONAL COMPONENT
N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL
N.B. D = CHRONOLOGICAL SAMPLE DATE

SUMMARY STATISTICS FOR DE-SEASONALIZED, DE-TRENDED RESIDUALS *****

| | |
|---|---|
| OBS | 212 |
| A-MEAN | -.0328 |
| STD DEV | .2148 |
| MAX | .3833 |
| Q90 | .1373 |
| Q75 | .0746 |
| Q50 | -.0006 |
| Q25 | -.0760 |
| Q10 | -.1889 |
| MIN | -1.1360 |

RANGE STATS

| | |
|---|---|
| MAX-MIN | 1.5192 |
| Q90-Q10 | .3263 |
| Q75-Q25 | .1506 |

TUKEY FENCES (STANDARDIZED) FOR OUTLIER SCAN

| | | |
|---|---|---|
| UPPER OUTER FENCE = | .5264 | 6.9948 |
| UPPER INNER FENCE = | .3005 | 3.9932 |
| LOWER INNER FENCE = | -.3018 | -4.0109 |
| LOWER OUTER FENCE = | -.5277 | -7.0125 |

!!! OUTLIERS IN DE-TRENDED & DE-SEASONALIZED RESIDUALS !!!

| | DATE | TIME | RAW DATUM | LOG10 | RESIDUAL | STANDARDIZED | |
|---|---|---|---|---|---|---|---|
| RESIDUAL | 741114 | 1200 | 151.000 | 2.179 | .383 | 5.093 | IS OUT |
| RESIDUAL | 760511 | 1200 | 126.000 | 2.100 | .313 | 4.155 | IS OUT |
| RESIDUAL | 760607 | 1200 | 19.000 | 1.279 | -.446 | -5.932 | IS OUT |
| RESIDUAL | 780228 | 1200 | 20.000 | 1.301 | -.788 | -10.477 | IS FAR OUT |
| RESIDUAL | 800730 | 1055 | 14.000 | 1.146 | -.660 | -8.768 | IS FAR OUT |
| RESIDUAL | 820914 | 830 | 24.000 | 1.380 | -.366 | -4.862 | IS OUT |
| RESIDUAL | 821018 | 1130 | 20.000 | 1.301 | -.452 | -6.011 | IS ALMOST FAR OUT |
| RESIDUAL | 840306 | 905 | 50.000 | 1.699 | .300 | 3.981 | IS ALMOST OUT |
| RESIDUAL | 840828 | 955 | 5.000 | .699 | -.481 | -6.391 | IS ALMOST FAR OUT |
| RESIDUAL | 850924 | 950 | 12.000 | 1.079 | -.300 | -3.989 | IS ALMOST OUT |
| RESIDUAL | 870512 | 1200 | 76.000 | 1.881 | .319 | 4.240 | IS OUT |
| RESIDUAL | 870805 | 1000 | 13.000 | 1.114 | -.315 | -4.180 | IS OUT |
| RESIDUAL | 870901 | 1100 | 2.000 | .301 | -1.136 | -15.096 | IS FAR OUT |
| RESIDUAL | 870928 | 1010 | 4.000 | .602 | -.848 | -11.263 | IS FAR OUT |
| RESIDUAL | 880803 | 1020 | 5.000 | .699 | -.649 | -8.622 | IS FAR OUT |
| RESIDUAL | 880830 | 1030 | 2.000 | .301 | -1.053 | -13.993 | IS FAR OUT |
| RESIDUAL | 880926 | 1000 | 2.000 | .301 | -1.067 | -14.175 | IS FAR OUT |
| RESIDUAL | 890516 | 900 | 57.000 | 1.756 | .302 | 4.007 | IS OUT |
| RESIDUAL | 890613 | 850 | 51.000 | 1.708 | .314 | 4.167 | IS OUT |
| RESIDUAL | 890925 | 1055 | 4.000 | .602 | -.751 | -9.974 | IS FAR OUT |
| RESIDUAL | 891128 | 1116 | 10.000 | 1.000 | -.383 | -5.090 | IS OUT |

THERE ARE:
2 ALMOST-OUT DATA
9 OUT DATA
2 ALMOST FAR OUT DATA
8 FAR OUT DATA
21 TOTAL OUTLIERS IDENTIFIED

N.B. OUTLIERS ARE STANDARDIZED BY DIVISION BY THE MEDIAN ABSOLUTE RESIDUAL = S

FOR NORMAL DISTRIBUTIONS:
OUTER FENCES ARE ABOUT 7S OR 4.7 X STANDARD DEVIATION
INNER FENCES ARE ABOUT 4S OR 2.7 X STANDARD DEVIATION
ALSO 6S OR 4 X STANDARD DEVIATION (ALMOST FAR OUT)
& 3.75S OR 2.5 X STANDARD DEVIATION (ALMOST OUT)
ARE STANDARDIZED VALUES USE BY OTHER AUTHORS

N.B. TUKEY FENCES ARE DEFINED USING 1/2 THE INTERQUARTILE RANGE
OF RESIDUALS WHICH IS NOT EXACTLY IDENTICAL TO S
EXCEPT FOR NORMAL DISTRIBUTIONS BUT IS ALSO A COMMON
ROBUST MEASURE OF DISPERSION

```
        PROGRAM TR1

 REPORT DATA FILE = D:\DILDO\TR1\CRno3
 INPUT DATA FILE = D:\DILDO\DAT2\NO3.dat
 VARIABLE = NO3
(5x,i7,i5,f9.3)
 START YEAR = 720000   END YEAR = 920000
  ZERO OPTION CODE = 1
  NUMERIC MISSING VALUE CODE =    -9.00000000
  MULTIPLIER/DIVISOR CODE = 1
  MULTIPLIER/DIVISOR =       .00000000
  INPUT DATA ARE LEFT AS IS
  LOGG CODE  =   0
  RUNS =   2
  FILTER  =  5
  SAMPLING INTERVAL CODE =  1
  MINIMUM SMOOTHING BASED ON MEDIAN SAMPLING INTERVAL
  UNIT SMOOTHING INTERVAL =   1.00 yr
  PLOT CODES:
 PLOT INDICATOR = 0
 PLOT DENSITY = 0
 LOG PLOT TYPE  = 0
 PLOT TYPE CODE = L  (i.e. LONG/SHORT)
 PLOT HEADER 1 = Nitrate ug/L
 PLOT HEADER 2 = CLEARWATER L
 RUN DATE HEADER = JUN 13 1991
 Y-AXIS TITLE = Concentration ug/L
 REPORT FILE CODE = SR  (i.e. SHORT REPORT)
 OUTPUT FILE CODE = 0
 NO OUTPUT FILES REQUESTED
 FIRST LINE OF INPUT DATA FILE:
  PARAMETER = NO3      SITE = CLEARWATER L

  QUALITY DATA FOR PARM = NO3     READ & STORED
   213 RECORDS WERE READ

  THE FOLLOWING    3 INPUT DATA WERE IDENITIFIED AS ZERO'S:
     870901 1100      .000  ZERO
     880830 1030      .000  ZERO
     880926 1000      .000  ZERO

  SMALLEST NON-ZERO POSITIVE VALUE =     4.00000

  ZERO OPTION CODE = 1
  SETTING 0'S TO 1/2 SMALLEST POSITIVE NON-ZERO VALUE =     2.00000

*******************************************************************************
    BEGINNING RUN  1

    UNIT SMOOTHING INTERVAL (USI) =      1.00000 yr
    LENGTH OF RECORD          =     16.60808 yr
    MEAN SAMPLING INTERVAL    =       .07834 yr
    MEAN SAMPLING RATE/USI    =     12.76487 #/USI
    MEDIAN SAMPLING INTERVAL =        .07382 yr
    MEDIAN SAMPLING RATE/USI =     13.54609 #/USI
    MINIMUM SMOOTHING         =    14 OBSERVATIONS


         TREND SUMMARY:

    GRAND TIME-WEIGHTED MEAN TREND =     53.354

    SCAN SMOOTH FOR LOCAL EXTREMA INCLUDING END POINTS
```

```
LOCAL MAX =   110.130 @  73.489 = JUN 28 1973
LOCAL MIN =    66.406 @  74.788 = OCT 15 1974
LOCAL MAX =    67.009 @  74.870 = NOV 14 1974
LOCAL MIN =    55.109 @  76.818 = OCT 26 1976
LOCAL MAX =    87.733 @  78.656 = AUG 28 1978
LOCAL MIN =    74.436 @  79.793 = OCT 17 1979
LOCAL MAX =    75.176 @  80.362 = MAY 12 1980
LOCAL MIN =    24.383 @  84.657 = AUG 28 1984
LOCAL MAX =    31.637 @  85.387 = MAY 22 1985
LOCAL MIN =    29.892 @  85.807 = OCT 22 1985
LOCAL MAX =    39.577 @  87.113 = FEB 11 1987
LOCAL MIN =    26.131 @  88.736 = SEP 26 1988
LOCAL MAX =    28.713 @  89.223 = MAR 23 1989
LOCAL MIN =    10.904 @  90.097 = FEB  5 1990

MAX TREND =   110.130 @  73.489 = JUN 28 1973
MIN TREND =    10.904 @  90.097 = FEB  5 1990
TREND RANGE =   99.225
```

ANNUAL TIME-WEIGHTED MEAN TREND SUMMARY

| | YEAR | N | MEAN | YR-TO-YR | DIFF |
|---|---|---|---|---|---|
| 1 | 73 | 16 | 104.811 | 74-73 | -36.470 |
| 2 | 74 | 21 | 68.342 | 75-74 | -2.597 |
| 3 | 75 | 10 | 65.744 | 76-75 | -8.198 |
| 4 | 76 | 13 | 57.546 | 77-76 | 8.456 |
| 5 | 77 | 10 | 66.003 | 78-77 | 18.847 |
| 6 | 78 | 14 | 84.850 | 79-78 | -7.236 |
| 7 | 79 | 19 | 77.613 | 80-79 | -3.276 |
| 8 | 80 | 12 | 74.338 | 81-80 | -7.771 |
| 9 | 81 | 7 | 66.567 | 82-81 | -13.120 |
| 10 | 82 | 7 | 53.447 | 83-82 | -16.240 |
| 11 | 83 | 10 | 37.207 | 84-83 | -10.984 |
| 12 | 84 | 13 | 26.223 | 85-84 | 4.390 |
| 13 | 85 | 12 | 30.614 | 86-85 | 4.354 |
| 14 | 86 | 12 | 34.968 | 87-86 | -.074 |
| 15 | 87 | 13 | 34.894 | 88-87 | -7.294 |
| 16 | 88 | 13 | 27.600 | 89-88 | -3.480 |
| 17 | 89 | 10 | 24.120 | 90-89 | -13.216 |
| 18 | 90 | 1 | 10.904 | | |

```
MEAN TREND    =    52.544

MAXIMUM TREND =   104.811 @ 1973
MINIMUM TREND =    10.904 @ 1990
TREND RANGE   =    93.907

SPEARMANS RHO TREND TEST (2-SIDED) ON FINAL ANNUAL TRENDS
FOR YEARS WITH WITH 5 OR MORE OBSERVATIONS
IF RHO < 0 THEN MET TREND IS DOWN
IF RHO > 0 THEN MET TREND IS UP
RHO =  -.82108   APPROXX p =  .000054   dof =  15


$$$ APPROXIMATE SOURCES OF VARIATION $$$

VARIATION DUE TO TREND    =  62.67897 %
VARIATION DUE TO NOISE    =  37.32103 %
APPROXIMATE RSQ           =  62.67897 %


MEAN X      =    55.469480     MEAN E      =    -1.075196
```

STD DEV X = 30.035590 STD DEV E = 18.363080

ROBUST SUMMARY STATISTICS:

| | | | |
|---|---|---|---|
| MEDIAN X | = 51.000000 | MEDIAN E | = -1.289604 |
| MEDIAN \|X\| | = 51.000000 | MEDIAN \|E\| | = 9.891596 |
| MAD X | = 21.000000 | MAD E | = 9.116722 |
| .5 X IQR | = 20.812500 | .5 E IQR | = 9.372667 |
| R STD DEV | = 31.134650 | R STD DEV | = 13.516470 |

N.B. X = RAW DATA
N.B. T = TREND COMPONENT
N.B. E = RESIDUAL = X-T

SUMMARY STATISTICS FOR RESIDUALS *****

| | |
|---|---|
| OBS | 213 |
| A-MEAN | -1.0752 |
| STD DEV | 18.3631 |
| MAX | 83.9915 |
| Q90 | 19.8664 |
| Q75 | 7.0706 |
| Q50 | -1.2896 |
| Q25 | -11.6748 |
| Q10 | -22.7806 |
| MIN | -61.2132 |
| RANGE STATS | |
| MAX-MIN | 145.2046 |
| Q90-Q10 | 42.6469 |
| Q75-Q25 | 18.7453 |

TUKEY FENCES (STANDARDIZED) FOR OUTLIER SCAN

| | | |
|---|---|---|
| UPPER OUTER FENCE = | 63.3066 | 6.4000 |
| UPPER INNER FENCE = | 35.1886 | 3.5574 |
| LOWER INNER FENCE = | -39.7928 | -4.0229 |
| LOWER OUTER FENCE = | -67.9108 | -6.8655 |

!!! OUTLIERS IN RESIDUALS !!!

| | DATE | TIME | RAW DATUM | RESIDUAL | STANDARDIZED | |
|---|---|---|---|---|---|---|
| RESIDUAL | 730809 | 1200 | 162.000 | 57.019 | 5.764 | IS OUT |
| RESIDUAL | 741114 | 1200 | 151.000 | 83.991 | 8.491 | IS FAR OUT |
| RESIDUAL | 760429 | 1200 | 96.000 | 36.050 | 3.644 | IS OUT |
| RESIDUAL | 760511 | 1200 | 126.000 | 66.579 | 6.731 | IS FAR OUT |
| RESIDUAL | 760607 | 1200 | 19.000 | -39.230 | -3.966 | IS ALMOST OUT |
| RESIDUAL | 780228 | 1200 | 20.000 | -61.213 | -6.188 | IS ALMOST FAR OUT |
| RESIDUAL | 790219 | 1200 | 121.000 | 38.247 | 3.867 | IS OUT |
| RESIDUAL | 790309 | 1200 | 128.000 | 46.090 | 4.659 | IS OUT |
| RESIDUAL | 790710 | 930 | 35.000 | -40.745 | -4.119 | IS OUT |
| RESIDUAL | 800730 | 1055 | 14.000 | -60.295 | -6.096 | IS ALMOST FAR OUT |
| RESIDUAL | 870512 | 1200 | 76.000 | 38.788 | 3.921 | IS OUT |

THERE ARE: 1 ALMOST OUT DATA
6 OUT DATA
2 ALMOST FAR OUT DATA

2 FAR OUT DATA
11 TOTAL OUTLIERS IDENTIFIED

N.B. OUTLIERS ARE STANDARDIZED BY DIVISION BY THE MEDIAN ABSOLUTE RESIDUAL = S

FOR NORMAL DISTRIBUTIONS:
OUTER FENCES ARE ABOUT 7S OR 4.7 X STANDARD DEVIATION
INNER FENCES ARE ABOUT 4S OR 2.7 X STANDARD DEVIATION
ALSO 6S OR 4 X STANDARD DEVIATION (ALMOST FAR OUT)
& 3.75S OR 2.5 X STANDARD DEVIATION (ALMOST OUT)
ARE STANDARDIZED VALUES USE BY OTHER AUTHORS

N.B. TUKEY FENCES ARE DEFINED USING 1/2 THE INTERQUARTILE RANGE
OF RESIDUALS WHICH IS NOT EXACTLY IDENTICAL TO S
EXCEPT FOR NORMAL DISTRIBUTIONS BUT IS ALSO A COMMON
ROBUST MEASURE OF DISPERSION

**********

```
        PROGRAM TR1

 REPORT DATA FILE = D:\DILDO\TR1\CRno3L
 INPUT DATA FILE = D:\DILDO\DAT2\NO3.dat
 VARIABLE = NO3
(5X,i7,i5,f9.3)
 START YEAR = 720000   END YEAR = 920000
  ZERO OPTION CODE = 1
  NUMERIC MISSING VALUE CODE =    -9.00000000
  MULTIPLIER/DIVISOR CODE = 1
  MULTIPLIER/DIVISOR =       .00000000
  INPUT DATA ARE LEFT AS IS
  LOGG CODE  =   1
  RUNS =   1
  FILTER  =  5
  SAMPLING INTERVAL CODE =  1
  MINIMUM SMOOTHING BASED ON MEDIAN SAMPLING INTERVAL
  UNIT SMOOTHING INTERVAL =   1.00 yr
  PLOT CODES:
 PLOT INDICATOR = 0
 PLOT DENSITY = D
 LOG PLOT TYPE  = 0
 PLOT TYPE CODE = L  (i.e. LONG/SHORT)
 PLOT HEADER 1 = Nitrate ug/L
 PLOT HEADER 2 = CLEARWATER L
 RUN DATE HEADER = JUN 13 1991
 Y-AXIS TITLE = Log10 Concentration ug/L
 REPORT FILE CODE = SR  (i.e. SHORT REPORT)
 OUTPUT FILE CODE = 0
 NO OUTPUT FILES REQUESTED
 FIRST LINE OF INPUT DATA FILE:
  PARAMETER = NO3      SITE = CLEARWATER L

  QUALITY DATA FOR PARM = NO3     READ & STORED
   213 RECORDS WERE READ

  THE FOLLOWING    3 INPUT DATA WERE IDENITIFIED AS ZERO'S:
     870901 1100       .000  ZERO
     880830 1030       .000  ZERO
     880926 1000       .000  ZERO

  SMALLEST NON-ZERO POSITIVE VALUE =     4.00000

  ZERO OPTION CODE = 1
  SETTING 0'S TO 1/2 SMALLEST POSITIVE NON-ZERO VALUE =     2.00000


  DOING LOG10 TRANSFORM OF DATA SERIES

*******************************************************************************
    BEGINNING RUN  1

    UNIT SMOOTHING INTERVAL (USI) =      1.00000 yr
    LENGTH OF RECORD          =     16.60808 yr
    MEAN SAMPLING INTERVAL    =       .07834 yr
    MEAN SAMPLING RATE/USI    =     12.76487 #/USI
    MEDIAN SAMPLING INTERVAL  =       .07382 yr
    MEDIAN SAMPLING RATE/USI  =     13.54609 #/USI
    MINIMUM SMOOTHING         =    14 OBSERVATIONS


         TREND SUMMARY:
```

GRAND TIME-WEIGHTED MEAN TREND = 1.676 LOG UNITS
47.409 NATURAL UNITS

SCAN SMOOTH FOR LOCAL EXTREMA INCLUDING END POINTS

LOCAL MAX = 113.886 @ 73.489 = JUN 28 1973
LOCAL MIN = 66.730 @ 74.368 = MAY 15 1974
LOCAL MAX = 71.272 @ 74.870 = NOV 14 1974
LOCAL MIN = 54.671 @ 76.818 = OCT 26 1976
LOCAL MAX = 88.183 @ 78.656 = AUG 28 1978
LOCAL MIN = 71.584 @ 79.637 = AUG 21 1979
LOCAL MAX = 73.838 @ 80.362 = MAY 12 1980
LOCAL MIN = 22.910 @ 84.507 = JUL 4 1984
LOCAL MAX = 29.240 @ 85.483 = JUN 26 1985
LOCAL MIN = 28.836 @ 85.807 = OCT 22 1985
LOCAL MAX = 38.663 @ 87.113 = FEB 11 1987
LOCAL MIN = 11.456 @ 90.097 = FEB 5 1990

MAX TREND = 113.886 @ 73.489 = JUN 28 1973
MIN TREND = 11.456 @ 90.097 = FEB 5 1990
TREND RANGE = 102.430

ANNUAL TIME-WEIGHTED MEAN TREND SUMMARY ON LOG10 SER IES

| | YEAR | N | MEAN | YR-TO-YR | DIFF |
|---|---|---|---|---|---|
| 1 | 73 | 16 | 2.027 | 74-73 | -.194 |
| 2 | 74 | 21 | 1.833 | 75-74 | -.004 |
| 3 | 75 | 10 | 1.828 | 76-75 | -.065 |
| 4 | 76 | 13 | 1.764 | 77-76 | .046 |
| 5 | 77 | 10 | 1.809 | 78-77 | .119 |
| 6 | 78 | 14 | 1.928 | 79-78 | -.050 |
| 7 | 79 | 19 | 1.878 | 80-79 | -.014 |
| 8 | 80 | 12 | 1.864 | 81-80 | -.042 |
| 9 | 81 | 7 | 1.822 | 82-81 | -.088 |
| 10 | 82 | 7 | 1.734 | 83-82 | -.178 |
| 11 | 83 | 10 | 1.557 | 84-83 | -.172 |
| 12 | 84 | 13 | 1.385 | 85-84 | .072 |
| 13 | 85 | 12 | 1.457 | 86-85 | .072 |
| 14 | 86 | 12 | 1.528 | 87-86 | .015 |
| 15 | 87 | 13 | 1.543 | 88-87 | -.096 |
| 16 | 88 | 13 | 1.447 | 89-88 | -.141 |
| 17 | 89 | 10 | 1.306 | 90-89 | -.247 |
| 18 | 90 | 1 | 1.059 | | |

MEAN TREND = 1.654

MAXIMUM TREND = 2.027 @ 1973
MINIMUM TREND = 1.059 @ 1990
TREND RANGE = .968

ANNUAL T-W MEAN TREND SUMMARY ON NATURAL SERIES

| | YEAR | N | MEAN | YR-TO-YR | DIFF |
|---|---|---|---|---|---|
| 1 | 73 | 16 | 106.395 | 74-73 | -38.366 |
| 2 | 74 | 21 | 68.029 | 75-74 | -.694 |
| 3 | 75 | 10 | 67.335 | 76-75 | -9.310 |
| 4 | 76 | 13 | 58.025 | 77-76 | 6.427 |
| 5 | 77 | 10 | 64.452 | 78-77 | 20.246 |
| 6 | 78 | 14 | 84.698 | 79-78 | -9.161 |
| 7 | 79 | 19 | 75.537 | 80-79 | -2.368 |
| 8 | 80 | 12 | 73.169 | 81-80 | -6.767 |
| 9 | 81 | 7 | 66.402 | 82-81 | -12.142 |
| 10 | 82 | 7 | 54.260 | 83-82 | -18.238 |

| | | | | | |
|---|---|---|---|---|---|
| 11 | 83 | 10 | 36.022 | 84-83 | -11.775 |
| 12 | 84 | 13 | 24.247 | 85-84 | 4.382 |
| 13 | 85 | 12 | 28.629 | 86-85 | 5.124 |
| 14 | 86 | 12 | 33.754 | 87-86 | 1.172 |
| 15 | 87 | 13 | 34.926 | 88-87 | -6.904 |
| 16 | 88 | 13 | 28.022 | 89-88 | -7.783 |
| 17 | 89 | 10 | 20.239 | 90-89 | -8.783 |
| 18 | 90 | 1 | 20.239 | | |

G-MEAN TREND = 47.409

MAXIMUM TREND = 106.395 @ 1973
MINIMUM TREND = 11.456 @ 1990
TREND RANGE = 94.939

SPEARMANS RHO TREND TEST (2-SIDED) ON FINAL ANNUAL TRENDS
FOR YEARS WITH WITH 5 OR MORE OBSERVATIONS
IF RHO < 0 THEN NET TREND IS DOWN
IF RHO > 0 THEN NET TREND IS UP
RHO = -.83824 APPROXX p = .000026 dof = 15

$$$ APPROXIMATE SOURCES OF VARIATION $$$

VARIATION DUE TO TREND = 48.30162 %
VARIATION DUE TO NOISE = 51.69838 %
APPROXIMATE RSQ = 48.30162 %

| | | | | | |
|---|---|---|---|---|---|
| MEAN X | = | 1.655521 | MEAN E | = | -.045038 |
| STD DEV X | = | .326339 | STD DEV E | = | .230726 |

ROBUST SUMMARY STATISTICS:

| | | | | | |
|---|---|---|---|---|---|
| MEDIAN X | = | 1.707570 | MEDIAN E | = | -.007329 |
| MEDIAN \|X\| | = | 1.707570 | MEDIAN \|E\| | = | .086720 |
| MAD X | = | .167491 | MAD E | = | .087448 |
| .5 X IQR | = | .175905 | .5 E IQR | = | .095139 |
| R STD DEV | = | .248323 | R STD DEV | = | .129650 |

N.B. X = RAW DATA
N.B. T = TREND COMPONENT
N.B. E = RESIDUAL = X-T

SUMMARY STATISTICS FOR RESIDUALS *****

| | |
|---|---|
| OBS | 213 |
| A-MEAN | -.0450 |
| STD DEV | .2307 |
| MAX | .3836 |
| Q90 | .1712 |
| Q75 | .0718 |
| Q50 | -.0073 |
| Q25 | -.1185 |
| Q10 | -.2303 |
| MIN | -1.2188 |
| RANGE STATS | |
| MAX-MIN | 1.6023 |

| | |
|---|---|
| Q90-Q10 | .4015 |
| Q75-Q25 | .1903 |

TUKEY FENCES (STANDARDIZED) FOR OUTLIER SCAN

| | | |
|---|---|---|
| UPPER OUTER FENCE = | .6426 | 7.4104 |
| UPPER INNER FENCE = | .3572 | 4.1192 |
| LOWER INNER FENCE = | -.4039 | -4.6574 |
| LOWER OUTER FENCE = | -.6893 | -7.9486 |

!!! OUTLIERS IN RESIDUALS !!!

| | DATE | TIME | RAW DATUM | LOG10 | RESIDUAL | STANDARDIZED | |
|---|---|---|---|---|---|---|---|
| RESIDUAL | 741114 | 1200 | 151.000 | 2.179 | .326 | 3.760 | IS ALMOST OUT |
| RESIDUAL | 760607 | 1200 | 19.000 | 1.279 | -.494 | -5.691 | IS OUT |
| RESIDUAL | 780228 | 1200 | 20.000 | 1.301 | -.604 | -6.971 | IS ALMOST FAR OUT |
| RESIDUAL | 800730 | 1055 | 14.000 | 1.146 | -.719 | -8.293 | IS FAR OUT |
| RESIDUAL | 820914 | 830 | 24.000 | 1.380 | -.334 | -3.853 | IS ALMOST OUT |
| RESIDUAL | 821018 | 1130 | 20.000 | 1.301 | -.396 | -4.570 | IS ALMOST OUT |
| RESIDUAL | 840828 | 955 | 5.000 | .699 | -.662 | -7.635 | IS ALMOST FAR OUT |
| RESIDUAL | 840925 | 920 | 10.000 | 1.000 | -.371 | -4.283 | IS ALMOST OUT |
| RESIDUAL | 850924 | 950 | 12.000 | 1.079 | -.382 | -4.407 | IS ALMOST OUT |
| RESIDUAL | 870805 | 1000 | 13.000 | 1.114 | -.418 | -4.820 | IS OUT |
| RESIDUAL | 870901 | 1100 | 2.000 | .301 | -1.219 | -14.054 | IS FAR OUT |
| RESIDUAL | 870928 | 1010 | 4.000 | .602 | -.908 | -10.467 | IS FAR OUT |
| RESIDUAL | 880706 | 1010 | 13.000 | 1.114 | -.330 | -3.805 | IS ALMOST OUT |
| RESIDUAL | 880803 | 1020 | 5.000 | .699 | -.740 | -8.528 | IS FAR OUT |
| RESIDUAL | 880830 | 1030 | 2.000 | .301 | -1.132 | -13.056 | IS FAR OUT |
| RESIDUAL | 880926 | 1000 | 2.000 | .301 | -1.128 | -13.002 | IS FAR OUT |
| RESIDUAL | 890516 | 900 | 57.000 | 1.756 | .384 | 4.423 | IS OUT |
| RESIDUAL | 890613 | 850 | 51.000 | 1.708 | .360 | 4.147 | IS OUT |
| RESIDUAL | 890925 | 1055 | 4.000 | .602 | -.627 | -7.225 | IS ALMOST FAR OUT |

THERE ARE:
- 6 ALMOST OUT DATA
- 4 OUT DATA
- 3 ALMOST FAR OUT DATA
- 6 FAR OUT DATA
- 19 TOTAL OUTLIERS IDENTIFIED

N.B. OUTLIERS ARE STANDARDIZED BY DIVISION BY THE MEDIAN ABSOLUTE RESIDUAL = S

FOR NORMAL DISTRIBUTIONS:
OUTER FENCES ARE ABOUT 7S OR 4.7 X STANDARD DEVIATION
INNER FENCES ARE ABOUT 4S OR 2.7 X STANDARD DEVIATION
ALSO 6S OR 4 X STANDARD DEVIATION (ALMOST FAR OUT)
& 3.75S OR 2.5 X STANDARD DEVIATION (ALMOST OUT)
ARE STANDARDIZED VALUES USE BY OTHER AUTHORS

N.B. TUKEY FENCES ARE DEFINED USING 1/2 THE INTERQUARTILE RANGE OF RESIDUALS WHICH IS NOT EXACTLY IDENTICAL TO S EXCEPT FOR NORMAL DISTRIBUTIONS BUT IS ALSO A COMMON ROBUST MEASURE OF DISPERSION

```
        PROGRAM TR2

 REPORT DATA FILE = D:\DILDO\TR2\CRno3
 INPUT DATA FILE = D:\DILDO\DAT2\NO3.dat
 VARIABLE = NO3
(5x,i7,i5,f9.3)
 START YEAR = 720000   END YEAR = 920000
  ZERO OPTION CODE = 1
  NUMERIC MISSING VALUE CODE =    -9.00000000
  MULTIPLIER/DIVISOR CODE = 1
  MULTIPLIER/DIVISOR =       .00000000
  INPUT DATA ARE LEFT AS IS
  LOGG CODE  =   0
  RUNS =   1
  ITERATIONS =   5
  FILTER =  5
  SAMPLING INTERVAL CODE =  1
  MINIMUM SMOOTHING BASED ON MEDIAN SAMPLING INTERVAL
  UNIT SMOOTHING INTERVAL =   1.00 yr
  MINIMUM N / WINDOW =  13
  MINIMUM SEASONAL SMOOTHING WINDOW = 1/  8 yr
  PLOT CODES:
 PLOT INDICATOR = 0
 PLOT DENSITY 1 = D
 LOG PLOT TYPE  = 0
 PLOT TYPE CODE = L  (i.e. LONG/SHORT)
 PLOT HEADER 1 = Nitrate ug/L
 PLOT HEADER 2 = CLEARWATER L
 RUN DATE HEADER = JUN 13 1991
 Y-AXIS TITLE = Concentration ug/L
 REPORT FILE CODE = SR  (i.e. SHORT REPORT)
 OUTPUT FILE CODE = 0
 NO OUTPUT FILES REQUESTED
 FIRST LINE OF INPUT DATA FILE:
  PARAMETER = NO3       SITE = CLEARWATER L

  QUALITY DATA FOR PARM = NO3      READ & STORED
   213 RECORDS WERE READ

  THE FOLLOWING    3 INPUT DATA WERE IDENITIFIED AS ZERO'S:
     870901 1100       .000  ZERO
     880830 1030       .000  ZERO
     880926 1000       .000  ZERO

  SMALLEST NON-ZERO POSITIVE VALUE =      4.00000

  ZERO OPTION CODE = 1
  SETTING 0'S TO 1/2 SMALLEST POSITIVE NON-ZERO VALUE =      2.00000

*******************************************************************************
    BEGINNING RUN  1

    UNIT SMOOTHING INTERVAL (USI) =       1.00000 yr
    LENGTH OF RECORD              =      16.60808 yr
    MEAN SAMPLING INTERVAL        =        .07834 yr
    MEAN SAMPLING RATE/USI        =      12.76487 #/USI
    MEDIAN SAMPLING INTERVAL      =        .07382 yr
    MEDIAN SAMPLING RATE/USI      =      13.54609 #/USI
    MINIMUM SMOOTHING             =     14 OBSERVATIONS


         TREND SUMMARY:
```

GRAND TIME-WEIGHTED MEAN TREND = 55.830

SCAN SMOOTH FOR LOCAL EXTREMA INCLUDING END POINTS

LOCAL MAX = 122.408 @ 73.489 = JUN 28 1973
LOCAL MIN = 74.368 @ 74.788 = OCT 15 1974
LOCAL MAX = 74.900 @ 74.870 = NOV 14 1974
LOCAL MIN = 58.842 @ 76.818 = OCT 26 1976
LOCAL MAX = 91.517 @ 78.656 = AUG 28 1978
LOCAL MIN = 76.350 @ 79.793 = OCT 17 1979
LOCAL MAX = 76.979 @ 80.362 = MAY 12 1980
LOCAL MIN = 25.084 @ 84.507 = JUL 4 1984
LOCAL MAX = 32.026 @ 85.387 = MAY 22 1985
LOCAL MIN = 30.215 @ 85.946 = DEC 12 1985
LOCAL MAX = 38.758 @ 86.960 = DEC 17 1986
LOCAL MIN = 25.412 @ 88.736 = SEP 26 1988
LOCAL MAX = 27.214 @ 89.223 = MAR 23 1989
LOCAL MIN = 11.397 @ 90.097 = FEB 5 1990

MAX TREND = 122.408 @ 73.489 = JUN 28 1973
MIN TREND = 11.397 @ 90.097 = FEB 5 1990
TREND RANGE = 111.011

ANNUAL TIME-WEIGHTED MEAN TREND SUMMARY

| | YEAR | N | ITER = 1 | ITER = 5 | DIFF | YR-TO-YR | DIFF |
|---|---|---|---|---|---|---|---|
| 1 | 73 | 16 | 104.811 | 115.881 | -11.070 | 74-73 | -39.032 |
| 2 | 74 | 21 | 68.342 | 76.849 | -8.507 | 75-74 | -6.731 |
| 3 | 75 | 10 | 65.744 | 70.118 | -4.373 | 76-75 | -7.964 |
| 4 | 76 | 13 | 57.546 | 62.154 | -4.607 | 77-76 | 6.461 |
| 5 | 77 | 10 | 66.003 | 68.615 | -2.612 | 78-77 | 19.653 |
| 6 | 78 | 14 | 84.850 | 88.268 | -3.418 | 79-78 | -8.404 |
| 7 | 79 | 19 | 77.613 | 79.864 | -2.251 | 80-79 | -3.115 |
| 8 | 80 | 12 | 74.338 | 76.750 | -2.412 | 81-80 | -4.585 |
| 9 | 81 | 7 | 66.567 | 72.165 | -5.598 | 82-81 | -14.698 |
| 10 | 82 | 7 | 53.447 | 57.466 | -4.019 | 83-82 | -19.504 |
| 11 | 83 | 10 | 37.207 | 37.962 | -.755 | 84-83 | -11.384 |
| 12 | 84 | 13 | 26.223 | 26.578 | -.355 | 85-84 | 4.423 |
| 13 | 85 | 12 | 30.614 | 31.001 | -.388 | 86-85 | 3.902 |
| 14 | 86 | 12 | 34.968 | 34.903 | .065 | 87-86 | -1.074 |
| 15 | 87 | 13 | 34.894 | 33.829 | 1.064 | 88-87 | -7.503 |
| 16 | 88 | 13 | 27.600 | 26.326 | 1.274 | 89-88 | -3.165 |
| 17 | 89 | 10 | 24.120 | 23.161 | .959 | 90-89 | -11.764 |
| 18 | 90 | 1 | 10.904 | 11.397 | -.493 | | |

MEAN TREND = 53.354 55.183

MAXIMUM TREND = 115.881 @ 1973
MINIMUM TREND = 11.397 @ 1990
TREND RANGE = 104.484

SPEARMANS RHO TREND TEST ON FINAL ANNUAL TRENDS
FOR YEARS WITH WITH 5 OR MORE OBSERVATIONS
IF RHO < 0 THEN MET TREND IS DOWN
IF RHO > 0 THEN MET TREND IS UP
RHO = -.85049 APPROXX p = .000015 dof = 15

$$$$$$$ SEASONAL SUMMARY $$$$$$

MEAN ANNUAL SEASONAL MEDIAN = 55.093070

```
        SCAN FOR SEASONAL EXTREMA

LOCAL MAX =   17.941220 @  .21  =  MAR 18
LOCAL MIN =  -13.935940 @  .62  =  AUG 15

SEASONAL MAX =   17.941220 @  .21  =  MAR 18
SEASONAL MIN =  -13.935940 @  .62  =  AUG 15
SEASONAL AMPLITUDE =   31.877160

 ITERATIONS =  5  COMPLETED


$$$  VARIATION IN DE-SEASONALIZED RESIDUALS  $$$

   VARIATION DUE TO TREND    =  73.10892 %
   VARIATION DUE TO NOISE    =  26.89107 %
   APPROXIMATE RSQ           =  73.10892 %


   MEAN X      =    59.081190     MEAN E      =     -.426665
   STD DEV X   =    30.953310     STD DEV E   =    16.134380


   ROBUST SUMMARY STATISTICS:

   MEDIAN X    =    55.701320     MEDIAN E    =     -.555134
   MEDIAN |X|  =    55.701320     MEDIAN |E|  =     7.743704
   MAD X       =    23.790020     MAD E       =     7.283405
   .5 X IQR    =    23.812310     .5 E IQR    =     7.597422
   R STD DEV   =    35.271130     R STD DEV   =    10.798390

   N.B. X = DE-SESEASONALIZED RESIDUAL
   N.B. T = TREND COMPONENT
   N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL



$$$  VARIATION IN DE-TRENDED RESIDUALS  $$$

   VARIATION DUE TO SEASONAL    =  27.61154 %
   VARIATION DUE TO NOISE       =  72.38847 %
   APPROXIMATE RSQ              =  27.61153 %


   MEAN X      =    51.791650     MEAN E      =      .310286
   STD DEV X   =    18.676070     STD DEV E   =    16.134380


   ROBUST SUMMARY STATISTICS:

   MEDIAN X    =    51.109500     MEDIAN E    =      .181818
   MEDIAN |X|  =    51.109500     MEDIAN |E|  =     7.465222
   MAD X       =     9.823494     MAD E       =     7.283405
   .5 X IQR    =     9.879261     .5 E IQR    =     7.597422
   R STD DEV   =    14.564330     R STD DEV   =    10.798390

   N.B. X = DE-TRENDED RESIDUAL
   N.B. S = SEASONAL COMPONENT
   N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL
```

$$$ TOTAL VARIATION IN SERIES: $$$

| | |
|---|---|
| VARIATION DUE TO TREND | = 59.83679 % |
| VARIATION DUE TO SEASON | = -7.66329 % |
| VARIATION DUE TO S+T | = 71.14390 % |
| VARIATION DUE TO NOISE | = 28.85609 % |
| APPROXIMATE RSQ | = 71.14391 % |

| | | | |
|---|---|---|---|
| MEAN X | = 55.469480 | MEAN E | = -.058188 |
| STD DEV X | = 30.035590 | STD DEV E | = 16.134380 |

ROBUST SUMMARY STATISTICS

| | | | |
|---|---|---|---|
| MEDIAN X | = 51.000000 | MEDIAN E | = -.186657 |
| MEDIAN \|X\| | = 51.000000 | MEDIAN \|E\| | = 7.436587 |
| MAD X | = 21.000000 | MAD E | = 7.283407 |
| .5 X IQR | = 20.812500 | .5 E IQR | = 7.597423 |
| R STD DEV | = 31.134650 | R STD DEV | = 10.798400 |

N.B. X = RAW DATA
N.B. T = TREND COMPONENT
N.B. S = SEASONAL COMPONENT
N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL
N.B. D = CHRONOLOGICAL SAMPLE DATE

SUMMARY STATISTICS FOR DE-SEASONALIZED, DE-TRENDED RESIDUALS *****

| | |
|---|---|
| OBS | 213 |
| A-MEAN | -.0582 |
| STD DEV | 16.1344 |
| MAX | 78.9488 |
| Q90 | 16.8818 |
| Q75 | 7.0805 |
| Q50 | -.1867 |
| Q25 | -8.1143 |
| Q10 | -17.0278 |
| MIN | -80.1881 |

RANGE STATS

| | |
|---|---|
| MAX-MIN | 159.1370 |
| Q90-Q10 | 33.9096 |
| Q75-Q25 | 15.1948 |

TUKEY FENCES (STANDARDIZED) FOR OUTLIER SCAN

| | | |
|---|---|---|
| UPPER OUTER FENCE = | 52.6651 | 7.0819 |
| UPPER INNER FENCE = | 29.8728 | 4.0170 |
| LOWER INNER FENCE = | -30.9066 | -4.1560 |
| LOWER OUTER FENCE = | -53.6988 | -7.2209 |

!!! OUTLIERS IN DE-TRENDED & DE-SEASONALIZED RESIDUALS !!!

| | DATE | TIME | RAW DATUM | RESIDUAL | STANDARDIZED | |
|---|---|---|---|---|---|---|
| RESIDUAL | 730809 | 1200 | 162.000 | 60.176 | 8.092 | IS FAR OUT |
| RESIDUAL | 741114 | 1200 | 151.000 | 78.949 | 10.616 | IS FAR OUT |
| RESIDUAL | 751014 | 1200 | 91.000 | 30.592 | 4.114 | IS OUT |

| | | | | | | |
|---|---|---|---|---|---|---|
| RESIDUAL | 760511 | 1200 | 126.000 | 59.248 | 7.967 | IS FAR OUT |
| RESIDUAL | 760607 | 1200 | 19.000 | -40.061 | -5.387 | IS OUT |
| RESIDUAL | 780228 | 1200 | 20.000 | -80.188 | -10.783 | IS FAR OUT |
| RESIDUAL | 790710 | 930 | 35.000 | -34.367 | -4.621 | IS OUT |
| RESIDUAL | 800730 | 1055 | 14.000 | -49.704 | -6.684 | IS ALMOST FAR OUT |
| RESIDUAL | 870512 | 1200 | 76.000 | 37.768 | 5.079 | IS OUT |
| RESIDUAL | 880607 | 1027 | 50.000 | 28.153 | 3.786 | IS ALMOST OUT |
| RESIDUAL | 890516 | 900 | 57.000 | 29.195 | 3.926 | IS ALMOST OUT |
| RESIDUAL | 890613 | 850 | 51.000 | 30.421 | 4.091 | IS OUT |

THERE ARE:
2 ALMOST OUT DATA
5 OUT DATA
1 ALMOST FAR OUT DATA
4 FAR OUT DATA
12 TOTAL OUTLIERS IDENTIFIED

N.B. OUTLIERS ARE STANDARDIZED BY DIVISION BY THE MEDIAN ABSOLUTE RESIDUAL = S

FOR NORMAL DISTRIBUTIONS:
OUTER FENCES ARE ABOUT 7S OR 4.7 X STANDARD DEVIATION
INNER FENCES ARE ABOUT 4S OR 2.7 X STANDARD DEVIATION
ALSO 6S OR 4 X STANDARD DEVIATION (ALMOST FAR OUT)
& 3.75S OR 2.5 X STANDARD DEVIATION (ALMOST OUT)
ARE STANDARDIZED VALUES USE BY OTHER AUTHORS

N.B. TUKEY FENCES ARE DEFINED USING 1/2 THE INTERQUARTILE RANGE
OF RESIDUALS WHICH IS NOT EXACTLY IDENTICAL TO S
EXCEPT FOR NORMAL DISTRIBUTIONS BUT IS ALSO A COMMON
ROBUST MEASURE OF DISPERSION

```
        PROGRAM TR2

 REPORT DATA FILE = D:\DILDO\TR2\CRno3L
 INPUT DATA FILE = D:\DILDO\DAT2\NO3.dat
 VARIABLE = NO3
(5X,i7,i5,f9.3)
 START YEAR = 720000   END YEAR = 920000
  ZERO OPTION CODE = 1
  NUMERIC MISSING VALUE CODE =    -9.00000000
  MULTIPLIER/DIVISOR CODE = 1
  MULTIPLIER/DIVISOR =       .00000000
  INPUT DATA ARE LEFT AS IS
  LOGG CODE  =   1
  RUNS =   1
  ITERATIONS =   5
  FILTER =  5
  SAMPLING INTERVAL CODE =  1
  MINIMUM SMOOTHING BASED ON MEDIAN SAMPLING INTERVAL
  UNIT SMOOTHING INTERVAL =    1.00 yr
  MINIMUM N / WINDOW =  13
  MINIMUM SEASONAL SMOOTHING WINDOW = 1/  8 yr
  PLOT CODES:
 PLOT INDICATOR = 0
 PLOT DENSITY 1 = 0
 LOG PLOT TYPE  = 0
 PLOT TYPE CODE = L  (i.e. LONG/SHORT)
 PLOT HEADER 1 = Nitrate ug/L
 PLOT HEADER 2 = CLEARWATER L
 RUN DATE HEADER = JUN 13 1991
 Y-AXIS TITLE = Log10 Concentration ug/L
 REPORT FILE CODE = SR  (i.e. SHORT REPORT)
 OUTPUT FILE CODE = 0
 NO OUTPUT FILES REQUESTED
 FIRST LINE OF INPUT DATA FILE:
  PARAMETER = NO3      SITE = CLEARWATER L

  QUALITY DATA FOR PARM = NO3     READ & STORED
   213 RECORDS WERE READ

  THE FOLLOWING    3 INPUT DATA WERE IDENITIFIED AS ZERO'S:
     870901 1100      .000  ZERO
     880830 1030      .000  ZERO
     880926 1000      .000  ZERO

  SMALLEST NON-ZERO POSITIVE VALUE =     4.00000

  ZERO OPTION CODE = 1
  SETTING 0'S TO 1/2 SMALLEST POSITIVE NON-ZERO VALUE =     2.00000


  DOING LOG10 TRANSFORM OF DATA SERIES

*******************************************************************************
    BEGINNING RUN  1

    UNIT SMOOTHING INTERVAL (USI) =      1.00000 yr
    LENGTH OF RECORD          =     16.60808 yr
    MEAN SAMPLING INTERVAL    =       .07834 yr
    MEAN SAMPLING RATE/USI    =     12.76487 #/USI
    MEDIAN SAMPLING INTERVAL =       .07382 yr
    MEDIAN SAMPLING RATE/USI =     13.54609 #/USI
    MINIMUM SMOOTHING         =    14 OBSERVATIONS
```

TREND SUMMARY:

GRAND TIME-WEIGHTED MEAN TREND = 1.691 LOG UNITS
49.068 NATURAL UNITS

SCAN SMOOTH FOR LOCAL EXTREMA INCLUDING END POINTS

LOCAL MAX = 148.124 @ 73.489 = JUN 28 1973
LOCAL MIN = 81.997 @ 74.522 = JUL 10 1974
LOCAL MAX = 85.250 @ 74.870 = NOV 14 1974
LOCAL MIN = 58.727 @ 76.818 = OCT 26 1976
LOCAL MAX = 94.778 @ 78.656 = AUG 28 1978
LOCAL MIN = 73.930 @ 79.637 = AUG 21 1979
LOCAL MAX = 77.887 @ 80.788 = OCT 15 1980
LOCAL MIN = 22.945 @ 84.657 = AUG 28 1984
LOCAL MAX = 29.177 @ 85.483 = JUN 26 1985
LOCAL MIN = 28.706 @ 85.807 = OCT 22 1985
LOCAL MAX = 37.119 @ 86.960 = DEC 17 1986
LOCAL MIN = 12.201 @ 90.097 = FEB 5 1990

MAX TREND = 148.124 @ 73.489 = JUN 28 1973
MIN TREND = 12.201 @ 90.097 = FEB 5 1990
TREND RANGE = 135.923

ANNUAL TIME-WEIGHTED MEAN TREND SUMMARY ON LOG10 SER IES

| | YEAR | N | ITER = 1 | ITER = 5 | DIFF | YR-TO-YR | DIFF |
|---|---|---|---|---|---|---|---|
| 1 | 73 | 16 | 2.027 | 2.131 | -.105 | 74-73 | -.213 |
| 2 | 74 | 21 | 1.833 | 1.918 | -.085 | 75-74 | -.040 |
| 3 | 75 | 10 | 1.828 | 1.879 | -.050 | 76-75 | -.079 |
| 4 | 76 | 13 | 1.764 | 1.800 | -.036 | 77-76 | .035 |
| 5 | 77 | 10 | 1.809 | 1.835 | -.026 | 78-77 | .121 |
| 6 | 78 | 14 | 1.928 | 1.956 | -.028 | 79-78 | -.068 |
| 7 | 79 | 19 | 1.878 | 1.888 | -.010 | 80-79 | -.001 |
| 8 | 80 | 12 | 1.864 | 1.887 | -.023 | 81-80 | -.018 |
| 9 | 81 | 7 | 1.822 | 1.869 | -.047 | 82-81 | -.122 |
| 10 | 82 | 7 | 1.734 | 1.746 | -.012 | 83-82 | -.192 |
| 11 | 83 | 10 | 1.557 | 1.555 | .002 | 84-83 | -.169 |
| 12 | 84 | 13 | 1.385 | 1.386 | -.001 | 85-84 | .068 |
| 13 | 85 | 12 | 1.457 | 1.454 | .003 | 86-85 | .071 |
| 14 | 86 | 12 | 1.528 | 1.525 | .004 | 87-86 | -.016 |
| 15 | 87 | 13 | 1.543 | 1.509 | .034 | 88-87 | -.118 |
| 16 | 88 | 13 | 1.447 | 1.391 | .056 | 89-88 | -.111 |
| 17 | 89 | 10 | 1.306 | 1.280 | .026 | 90-89 | -.194 |
| 18 | 90 | 1 | 1.059 | 1.086 | -.027 | | |

MEAN TREND = 1.676 1.672

MAXIMUM TREND = 2.131 @ 1973
MINIMUM TREND = 1.086 @ 1990
TREND RANGE = 1.045

ANNUAL T-W MEAN TREND SUMMARY ON NATURAL SERIES

| | YEAR | N | ITER = 1 | ITER = 5 | DIFF | YR-TO-YR | DIFF |
|---|---|---|---|---|---|---|---|
| 1 | 73 | 16 | 106.395 | 135.351 | -28.956 | 74-73 | -52.527 |
| 2 | 74 | 21 | 68.029 | 82.824 | -14.795 | 75-74 | -7.203 |
| 3 | 75 | 10 | 67.335 | 75.621 | -8.286 | 76-75 | -12.511 |
| 4 | 76 | 13 | 58.025 | 63.110 | -5.085 | 77-76 | 5.276 |
| 5 | 77 | 10 | 64.452 | 68.385 | -3.934 | 78-77 | 21.998 |
| 6 | 78 | 14 | 84.698 | 90.384 | -5.686 | 79-78 | -13.103 |
| 7 | 79 | 19 | 75.537 | 77.281 | -1.744 | 80-79 | -.202 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 8 | 80 | 12 | 73.169 | 77.079 | -3.910 | 81-80 | -3.150 |
| 9 | 81 | 7 | 66.402 | 73.929 | -7.527 | 82-81 | -18.160 |
| 10 | 82 | 7 | 54.260 | 55.769 | -1.509 | 83-82 | -19.905 |
| 11 | 83 | 10 | 36.022 | 35.864 | .158 | 84-83 | -11.537 |
| 12 | 84 | 13 | 24.247 | 24.328 | -.081 | 85-84 | 4.115 |
| 13 | 85 | 12 | 28.629 | 28.443 | .187 | 86-85 | 5.028 |
| 14 | 86 | 12 | 33.754 | 33.471 | .283 | 87-86 | -1.209 |
| 15 | 87 | 13 | 34.926 | 32.262 | 2.664 | 88-87 | -7.649 |
| 16 | 88 | 13 | 28.022 | 24.613 | 3.409 | 89-88 | -5.538 |
| 17 | 89 | 10 | 20.239 | 19.075 | 1.164 | 90-89 | -6.874 |
| 18 | 90 | 1 | 11.456 | 12.201 | -.744 | | |

G-MEAN TREND = 47.409 49.068

MAXIMUM TREND = 135.351 @ 1973
MINIMUM TREND = 12.201 @ 1990
TREND RANGE = 123.150

SPEARMANS RHO TREND TEST ON FINAL ANNUAL TRENDS
FOR YEARS WITH WITH 5 OR MORE OBSERVATIONS
IF RHO < 0 THEN MET TREND IS DOWN
IF RHO > 0 THEN MET TREND IS UP
RHO = -.86765 APPROXX p = .000006 dof = 15

$$$$$$$ SEASONAL SUMMARY $$$$$$

MEAN ANNUAL SEASONAL MEDIAN = 1.690376 LOG UNITS

G-MEAN ANNUAL SEASONAL MEDIAN = 49.020320 NATURAL UNITS

SCAN FOR SEASONAL EXTREMA

LOCAL MAX = .190614 @ .25 = APR 2
LOCAL MIN = -.131510 @ .62 = AUG 15

SEASONAL MAX = .190614 @ .25 = APR 2
SEASONAL MIN = -.131510 @ .62 = AUG 15
SEASONAL AMPLITUDE = .322124

IN NATURAL UNITS
AT G-MEAN ANNUAL SEASONAL MEDIAN = 49.020

SEASONAL MAX = 76.030840 @ .25 = APR 2
SEASONAL MIN = 36.213120 @ .62 = AUG 15
SEASONAL AMPLITUDE = 39.817720

AT MAXIMUM TREND = 135.351 IN 1973

SEASONAL MAX = 209.930000 @ .25 = APR 2
SEASONAL MIN = 99.988640 @ .62 = AUG 15
SEASONAL AMPLITUDE = 109.941400

AT MINIMUM TREND = 12.201 IN 1990

SEASONAL MAX = 18.923100 @ .25 = APR 2
SEASONAL MIN = 9.012978 @ .62 = AUG 15
SEASONAL AMPLITUDE = 9.910118

ITERATIONS = 5 COMPLETED

$$$ VARIATION IN DE-SEASONALIZED RESIDUALS $$$

| | |
|---|---|
| VARIATION DUE TO TREND | = 62.19750 % |
| VARIATION DUE TO NOISE | = 37.80250 % |
| APPROXIMATE RSQ | = 62.19750 % |

| | | | | | |
|---|---|---|---|---|---|
| MEAN X | = | 1.69145D | MEAN E | = | -.029336 |
| STD DEV X | = | .326927 | STD DEV E | = | .198845 |

ROBUST SUMMARY STATISTICS:

| | | | | | |
|---|---|---|---|---|---|
| MEDIAN X | = | 1.749538 | MEDIAN E | = | .003463 |
| MEDIAN \|X\| | = | 1.749538 | MEDIAN \|E\| | = | .077308 |
| MAD X | = | .191625 | MAD E | = | .079055 |
| .5 X IQR | = | .205832 | .5 E IQR | = | .079137 |
| R STD DEV | = | .284104 | R STD DEV | = | .117207 |

N.B. X = DE-SESEASONALIZED RESIDUAL
N.B. T = TREND COMPONENT
N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL

$$$ VARIATION IN DE-TRENDED RESIDUALS $$$

| | |
|---|---|
| VARIATION DUE TO SEASONAL | = 29.24598 % |
| VARIATION DUE TO NOISE | = 70.75402 % |
| APPROXIMATE RSQ | = 29.24598 % |

| | | | | | |
|---|---|---|---|---|---|
| MEAN X | = | 1.625531 | MEAN E | = | -.028916 |
| STD DEV X | = | .229881 | STD DEV E | = | .198845 |

ROBUST SUMMARY STATISTICS:

| | | | | | |
|---|---|---|---|---|---|
| MEDIAN X | = | 1.641399 | MEDIAN E | = | .003883 |
| MEDIAN \|X\| | = | 1.641399 | MEDIAN \|E\| | = | .077601 |
| MAD X | = | .104543 | MAD E | = | .079055 |
| .5 X IQR | = | .104333 | .5 E IQR | = | .079137 |
| R STD DEV | = | .154996 | R STD DEV | = | .117207 |

N.B. X = DE-TRENDED RESIDUAL
N.B. S = SEASONAL COMPONENT
N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL

$$$ TOTAL VARIATION IN SERIES: $$$

| | |
|---|---|
| VARIATION DUE TO TREND | = 47.00081 % |
| VARIATION DUE TO SEASON | = .78965 % |
| VARIATION DUE TO S+T | = 62.50757 % |
| VARIATION DUE TO NOISE | = 37.49243 % |
| APPROXIMATE RSQ | = 62.50757 % |

| | | | |
|---|---|---|---|
| MEAN X = | 1.655521 | MEAN E = | -.029126 |
| STD DEV X = | .326339 | STD DEV E = | .198845 |

ROBUST SUMMARY STATISTICS

| | | | |
|---|---|---|---|
| MEDIAN X = | 1.707570 | MEDIAN E = | .003673 |
| MEDIAN \|X\| = | 1.707570 | MEDIAN \|E\| = | .077391 |
| MAD X = | .167491 | MAD E = | .079055 |
| .5 X IQR = | .175905 | .5 E IQR = | .079137 |
| R STD DEV = | .248323 | R STD DEV = | .117207 |

N.B. X = RAW DATA
N.B. T = TREND COMPONENT
N.B. S = SEASONAL COMPONENT
N.B. E = DE-TRENDED, DE-SESEASONALIZED RESIDUAL
N.B. D = CHRONOLOGICAL SAMPLE DATE

SUMMARY STATISTICS FOR DE-SEASONALIZED, DE-TRENDED RESIDUALS *****

| | |
|---|---|
| OBS | 213 |
| A-MEAN | -.0291 |
| STD DEV | .1988 |
| MAX | .4433 |
| Q90 | .1354 |
| Q75 | .0676 |
| Q50 | .0037 |
| Q25 | -.0907 |
| Q10 | -.1682 |
| MIN | -1.0545 |

RANGE STATS

| | |
|---|---|
| MAX-MIN | 1.4978 |
| Q90-Q10 | .3035 |
| Q75-Q25 | .1583 |

TUKEY FENCES (STANDARDIZED) FOR OUTLIER SCAN

| | | |
|---|---|---|
| UPPER OUTER FENCE = | .5424 | 7.0084 |
| UPPER INNER FENCE = | .3050 | 3.9407 |
| LOWER INNER FENCE = | -.3281 | -4.2398 |
| LOWER OUTER FENCE = | -.5655 | -7.3075 |

!!! OUTLIERS IN DE-TRENDED & DE-SEASONALIZED RESIDUALS !!!

| | DATE | TIME | RAW DATUM | LOG10 | RESIDUAL | STANDARDIZED | |
|---|---|---|---|---|---|---|---|
| RESIDUAL | 760607 | 1200 | 19.000 | 1.279 | -.494 | -6.383 | IS ALMOST FAR OUT |
| RESIDUAL | 780228 | 1200 | 20.000 | 1.301 | -.760 | -9.815 | IS FAR OUT |
| RESIDUAL | 800730 | 1055 | 14.000 | 1.146 | -.624 | -8.065 | IS FAR OUT |
| RESIDUAL | 821018 | 1130 | 20.000 | 1.301 | -.316 | -4.083 | IS ALMOST OUT |
| RESIDUAL | 840828 | 955 | 5.000 | .699 | -.535 | -6.908 | IS ALMOST FAR OUT |
| RESIDUAL | 870512 | 1200 | 76.000 | 1.881 | .325 | 4.195 | IS OUT |
| RESIDUAL | 870901 | 1100 | 2.000 | .301 | -1.055 | -13.626 | IS FAR OUT |
| RESIDUAL | 870928 | 1010 | 4.000 | .602 | -.765 | -9.885 | IS FAR OUT |
| RESIDUAL | 880607 | 1027 | 50.000 | 1.699 | .349 | 4.508 | IS OUT |
| RESIDUAL | 880803 | 1020 | 5.000 | .699 | -.556 | -7.181 | IS ALMOST FAR OUT |
| RESIDUAL | 880830 | 1030 | 2.000 | .301 | -.949 | -12.258 | IS FAR OUT |
| RESIDUAL | 880926 | 1000 | 2.000 | .301 | -.968 | -12.509 | IS FAR OUT |
| RESIDUAL | 890516 | 900 | 57.000 | 1.756 | .414 | 5.353 | IS OUT |

| | | | | | | |
|---|---|---|---|---|---|---|
| RESIDUAL 890613 | 850 | 51.000 | 1.708 | .443 | 5.727 | IS OUT |
| RESIDUAL 890711 | 830 | 41.000 | 1.613 | .400 | 5.167 | IS OUT |
| RESIDUAL 890925 | 1055 | 4.000 | .602 | -.509 | -6.573 | IS ALMOST FAR OUT |

THERE ARE:
1 ALMOST OUT DATA
5 OUT DATA
4 ALMOST FAR OUT DATA
6 FAR OUT DATA
16 TOTAL OUTLIERS IDENTIFIED

N.B. OUTLIERS ARE STANDARDIZED BY DIVISION BY THE MEDIAN ABSOLUTE RESIDUAL = S

FOR NORMAL DISTRIBUTIONS:
OUTER FENCES ARE ABOUT 7S OR 4.7 X STANDARD DEVIATION
INNER FENCES ARE ABOUT 4S OR 2.7 X STANDARD DEVIATION
ALSO 6S OR 4 X STANDARD DEVIATION (ALMOST FAR OUT)
& 3.75S OR 2.5 X STANDARD DEVIATION (ALMOST OUT)
ARE STANDARDIZED VALUES USE BY OTHER AUTHORS

N.B. TUKEY FENCES ARE DEFINED USING 1/2 THE INTERQUARTILE RANGE
OF RESIDUALS WHICH IS NOT EXACTLY IDENTICAL TO S
EXCEPT FOR NORMAL DISTRIBUTIONS BUT IS ALSO A COMMON
ROBUST MEASURE OF DISPERSION